PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/19, 15/79, C07K 14/52, A61K 38/19, C12N 1/19, 5/10

A1

(11) International Publication Number: WO 96/17062

(43) International Publication Date: 6 June 1996 (06.06.96)

(21) International Application Number: PCT/US95/14929

(22) International Filing Date: 15 November 1995 (15.11.95)

(30) Priority Data:
08/346,999 30 November 1994 (30.11.94) US

(71) Applicant: ZYMOGENETICS, INC. [US/US]; 1201 Eastlake Avenue East, Seattle, WA 98102 (US).

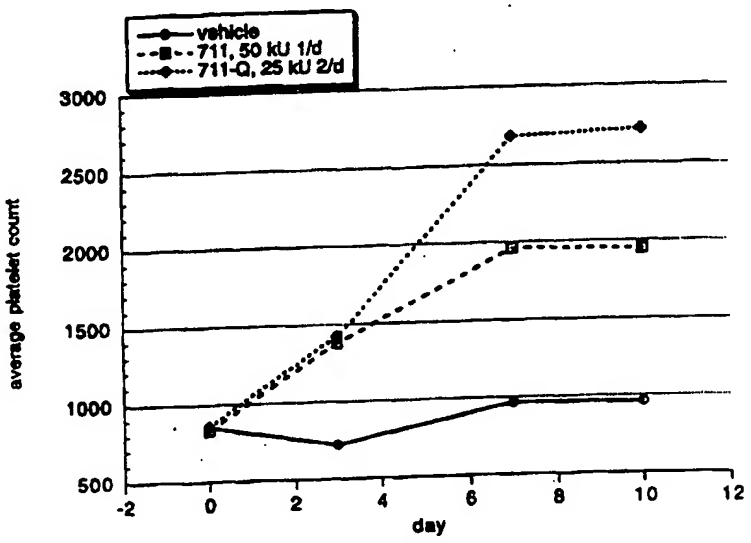
(72) Inventors: BUDDLE, Michele, M.; 32924 S.E. 46, Fall City, WA 98034 (US). DOWNEY, William; 6544 11th Avenue N.W., Seattle, WA 98117 (US). FOSTER, Donald, C.; 3002 N.E. 181st Street, Seattle, WA 98155 (US). FORSTROM, John, W.; 7559 30th Avenue N.E., Seattle, WA 98115 (US). LOK, Si; 806 N.W. 52nd, Seattle, WA 98107 (US). OSBORN, Sherri, G.; 8835 Ashworth Avenue North, Seattle, WA 98103 (US).

(74) Agent: PARKER, Gary, E.; ZymoGenetics, Inc., 1201 Eastlake Avenue East, Seattle, WA 98102 (US).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LT, LU, LV, MD, MG, MN, MW, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, LS, MW, SD, SZ, UG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: LOW MOLECULAR WEIGHT THROMBOPOIETIN



(57) Abstract

Isolated mammalian thrombopoietins are disclosed. The thrombopoietins include polypeptides characterized by an amino acid sequence selected from the group consisting of (a) a sequence of amino acids as shown in SEQ ID NO:2 having an amino terminus at Ser(45) and a carboxyl terminus between Ser(208) and Asn(216), inclusive; a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus between Arg(185) and Asn(193), inclusive; a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Arg(198); a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Phe(207); and allelic variants of (a)-(d). Also disclosed are isolated DNA molecules encoding these polypeptides as well as vectors, cells and methods that can be used for the production of the polypeptides.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

Description

LOW MOLECULAR WEIGHT THROMBOPOIETIN

Background of the Invention

Hematopoiesis is the process by which blood cells develop and differentiate from pluripotent stem cells in the bone marrow. This process involves a complex interplay of polypeptide growth factors (cytokines) acting via membrane-bound receptors on their target cells. Cytokine action results in cellular proliferation and differentiation, with response to a particular cytokine often being lineage-specific and/or stage-specific. Development of a single cell type, such as a platelet, from a stem cell may require the coordinated action of a plurality of cytokines acting in the proper sequence.

The known cytokines include the interleukins, such as IL-1, IL-2, IL-3, IL-6, IL-8, etc.; and the colony stimulating factors, such as G-CSF, M-CSF, GM-CSF, erythropoietin (EPO), etc. In general, the interleukins act as mediators of immune and inflammatory responses. The colony stimulating factors stimulate the proliferation of marrow-derived cells, activate mature leukocytes, and otherwise form an integral part of the host's response to inflammatory, infectious, and immunologic challenges.

Various cytokines have been developed as therapeutic agents. For example, erythropoietin, which stimulates the development of erythrocytes, is used in the treatment of anemia arising from renal failure. Several of the colony stimulating factors have been used in conjunction with cancer chemotherapy to speed the recovery of patients' immune systems. Interleukin-2, α-interferon and γ-interferon are used in the treatment of certain cancers. An activity that stimulates megakaryocytopoiesis and thrombocytopoiesis has been identified in body fluids

of thrombocytopenic animals and is referred to in the literature as "thrombopoietin" (recently reviewed by McDonald, Exp. Hematol. 16:201-205, 1988 and McDonald, Am. J. Ped. Hematol. Oncol. 14:8-21, 1992).

Recently, several groups have identified and/or cloned a protein that binds to the cellular MPL receptor and stimulates megakaryocytopoiesis and thrombocytopoiesis. See, de Sauvage et al., Nature 369:533-538, 1994; Lok et al., Nature 369:565-568, 1994; Kaushansky et al., Nature 369:568-571, 1994; Wendling et al., Nature 369:571-574, 1994; and Bartley et al., Cell 77:1117-1124, 1994. It has been proposed that this protein be termed thrombopoietin (Kaushansky et al., ibid.). Although this protein has been shown to stimulate platelet production *in vivo* (Kaushansky et al., ibid.), it appears to be subject to proteolysis and was isolated in heterogeneous or degraded form (Bartley et al., ibid.; de Sauvage et al., ibid.). Preparations of thrombopoietin reported in the scientific literature are therefore not well characterized as to composition and the relative activities of the various molecular species, although at least some of the proteolytic products are biologically active.

Proteolysis and heterogeneity are significant problems that can impede the development of new pharmaceutical agents. There thus remains a need in the art for reproducible preparations of thrombopoietin. There is a further need for preparations of thrombopoietin having high specific activity. There is a further need for homogeneous and well-characterized preparations of thrombopoietin. There is also a need for methods of making such preparations. The present invention fulfills these needs and provides other, related advantages.

Summary of the Invention

Within one aspect of the present invention there is provided an isolated mammalian thrombopoietin characterized by an amino acid sequence selected from the group consisting of (a) a sequence of amino acids as shown in SEQ ID NO:2 having an amino terminus at Ser(45) and a carboxyl terminus between Ser(208) and Asn(216), inclusive; (b) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus between Arg(185) and Asn(193), inclusive; (c) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Arg(198); (d) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Phe(207); and (e) allelic varaints of (a)-(d). Within one embodiment of the invention, the thrombopoietin is further characterized by a peptide backbone mass of 17,490 to 19,045 amu as determined by mass spectrometry, an absence of N-linked carbohydrate, a presence of O-linked carbohydrate, and being essentially free of thrombopoietin species having a molecular weight greater than 18,885 amu as determined by mass spectrometry. Within another embodiment of the invention, the thrombopoietin is a mixture of polypeptides having different carboxyl termini.

Within a related aspect, the present invention provides a composition of mammalian thrombopoietin consisting essentially of one or more polypeptides having an amino acid sequence consisting of (a) the sequence shown in SEQ ID NO:4 from an amino-terminus at Ser, residue number 22, to a carboxyl-terminus between Arg, residue number 185, and Asn, residue number 193, inclusive; and (b) allelic variants of (a). Within one embodiment, the composition is further characterized by an absence of N-linked carbohydrate and a presence of O-linked carbohydrate on the one or more polypeptides.

Within another aspect, the present invention provides a composition of mammalian thrombopoietin consisting essentially of one or more polypeptides having an amino acid sequence selected from the group consisting of (a) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Arg(198), (b) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Phe(207), and (c) allelic variants of (a) and (b).

Within another aspect of the present invention there is provided a pharmaceutical composition comprising a thrombopoietin polypeptide as disclosed above in combination with a pharmaceutically acceptable vehicle. Also provided is a method of stimulating platelet production in a mammal in need thereof comprising administering to the mammal a therapeutically effective amount of the composition.

Within an additional aspect of the invention, there is provided an isolated DNA molecule encoding a mammalian thrombopoietin as disclosed above. The DNA molecule may be used in the construction of an expression vector. The vector comprises a DNA segment encoding a thrombopoietin as disclosed above operably linked to one or more additional DNA segments that provide for its transcription. Also provided is a cultured eukaryotic cell containing the disclosed expression vector, such as a yeast cell or a cultured mammalian cell. The cultured cell may be used within methods for producing a mammalian thrombopoietin. These methods comprise the steps of culturing a eukaryotic cell containing an expression vector comprising a first DNA segment encoding a mammalian thrombopoietin as disclosed above, wherein the first DNA segment is operably linked to a second DNA segment encoding a secretory peptide and one or more additional DNA segments that provide for transcription of the first

and second DNA segments, and wherein the cell expresses the first and second DNA segments and the thrombopoietin is secreted from the cell and selectively recovered.

These and other aspects of the invention will become evident upon reference to the following detailed description and the attached drawings.

Brief Description of the Drawings

Fig. 1 illustrates the effects of a representative thrombopoietin composition of the present invention on platelet levels in mice.

Fig. 2 illustrates the dose-response relationship of platelet levels in mice treated with a representative thrombopoietin composition of the present invention.

Detailed Description of the Invention

Prior to describing the present invention in detail, it may be helpful to define certain terms used herein:

Allelic variant: An alternative form of a gene that arises through mutation, or an altered polypeptide encoded by the mutated gene. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequences.

cDNA: Complementary DNA, prepared by reverse transcription of a messenger RNA template, or a clone or amplified copy of such a molecule. Complementary DNA can be single-stranded or double-stranded.

Essentially free: At least 95% free of a specified contaminant. When applied to a protein, level of contamination is determined by Edman degradation and amino acid sequencing.

Expression vector: A DNA molecule, linear or circular, that comprises a segment encoding a polypeptide of interest operably linked to additional segments that

provide for its transcription. Such additional segments include promoter and terminator sequences, and may also include one or more origins of replication, one or more selectable markers, an enhancer, a polyadenylation signal, etc. Expression vectors are generally derived from plasmid or viral DNA, or may contain elements of both. The term "operably linked" indicates that the segments are arranged so that they function in concert for their intended purposes, e.g. transcription initiates in the promoter and proceeds through the coding segment to the terminator.

Gene: A segment of chromosomal DNA that encodes a polypeptide chain. A gene includes one or more regions encoding amino acids, which in some cases are interspersed with non-coding "intervening sequences" ("introns"), together with flanking, non-coding regions which provide for transcription of the coding sequence.

Isolated: When applied to the protein the term "isolated" indicates that the protein is found in a condition other than its native environment, such as apart from blood and animal tissue. In a preferred form, the isolated protein is substantially free of other proteins, particularly other proteins of animal origin. It is prefered to provide the proteins in a highly purified form, i.e. greater than 95% pure, more preferably greater than 99% pure. When applied to a polynucleotide molecule the term "isolated" indicates that the molecule is removed from its natural genetic milieu and is thus free of other extraneous or unwanted coding sequences, and is in a form suitable for use within genetically engineered protein production systems.

Peptide backbone mass: The molecular mass of a polypeptide or protien in the absence of glycosylation, generally determined by mass spectrometry of deglycosylated protein or calculated from amino acid sequence.

Promoter: The portion of a gene at which RNA polymerase binds and mRNA synthesis is initiated.

Secretory peptide: An amino acid sequence that acts to direct the secretion of a mature polypeptide or protein from a cell. Secretory peptides are characterized by a core of hydrophobic amino acids and are typically (but not exclusively) found at the amino termini of newly synthesized proteins. Very often the secretory peptide is cleaved from the mature protein during secretion. Such secretory peptides contain processing sites that allow cleavage of the secretory peptides from the mature proteins as they pass through the secretory pathway. A DNA sequence encoding a secretory peptide is referred to as a signal sequence, leader sequence, prepro sequence or pre sequence.

Thrombopoietin: Thrombopoietin (TPO) proteins are characterized by their ability to specifically bind to MPL receptor from the same species and to stimulate platelet production *in vivo*. In normal test animals, TPO is able to increase platelet levels by 100% or more within 10 days after beginning daily administration. The term "thrombopoietin polypeptide" encompasses full-length thrombopoietin molecules and biologically active portions thereof, that is fragments of a thrombopoietin that exhibit the qualitative biological activities of the intact molecule (receptor binding and *in vivo* stimulation of platelet production).

The present invention is based in part on the discovery of a class of defined and well-characterized, biologically active thrombopoietin polypeptides having unexpectedly high specific activity, members of which were initially obtained by fractionation of recombinant TPO. TPO compositions of the present invention exhibit an *in vitro* specific activity (units of activity per mole of protein) five to ten times that of conventional TPO preparations that comprise a mixture of glycosylated

polypeptide species ranging in size from about 30 kD to about 70 kD as determined by electrophoresis on SDS-polyacrylamide gels under reducing conditions. The TPO compositions of the present invention have also been found to be fully active *in vivo*.

A recombinant mouse TPO preparation that contained a significant amount of protein with a molecular weight in the 18-22 kD range was fractionated by a combination of affinity chromatography and anion exchange chromatography. This low molecular weight species was analyzed by mass spectrometry and found to have a heterogeneous carboxyl terminus, with the longest and most prevalent form ending at amino acid residue 216 of SEQ ID NO:2, and the shortest form ending at residue 208 of SEQ ID NO:2. The peptide backbone mass of the longest form, taking into consideration two disulfide bonds, was determined to be 18,449 atomic mass units (amu) by mass spectrometry. This low molecular weight TPO contained no N-linked carbohydrate as determined by the absence of a molecular weight shift after digestion with peptide-N-glycosidase F, which is specific for the hydrolysis of β-aspartyl-glycosylamine bonds between asparagine and the innermost N-acetylglucosamine of the glycan moiety; and by mass spectrometry, which indicated that the C-terminus of the polypeptide was upstream of the first N-linked glycosylation site of the full-length TPO molecule. The presence of O-linked carbohydrates on this TPO was determined by a shift in molecular weight after deglycosylation with O-glycosidase. Digestion with sialidase and O-glycosidase in combination with mass spectrometry indicated the presence of at least four different O-linked glycosylation sites with up to eight different glycoforms. The mass range of the carbohydrate structures was from 2260 to 3207. The largest form contained 4 HexNAc, 4 hexose, and 6 NeuAc (sialic acid). The peptide backbone mass of the shortest polypeptide in

this preparation (terminating at residue 208 of SEQ ID NO:2) was 17,492 amu.

The thrombopoietin compositions of the present invention may be a single polypeptide species or a mixture of polypeptides differing at their carboxyl termini but having a common amino terminus. A representative group of such polypeptides derived from mouse TPO includes molecules having a sequence of amino acids as shown in SEQ ID NO:2 from an amino-terminal Ser residue, amino acid residue no. 45, to a carboxyl terminus between residue no. 208 (Ser) and residue no. 216 (Asn), inclusive.

The present invention also provides related TPO polypeptides from species other than mouse. For example, there is provided a group of human TPO polypeptides, the amino acid sequences of which begin at Ser, residue 22 of SEQ ID NO:4 and terminate between residue 185 (Arg) and residue 198 (Arg), inclusive, of SEQ ID NO:4. Of particular interest are the human TPO polypeptides shown in Table 1 (with reference to SEQ ID NO:4), as well as intermediate forms, such as polypeptides having C-termini between Arg (185) and Asn (193).

Table 1

Ser (22)--Arg (185)
Ser (22)--Asn (193)
Ser (22)--Arg (198)

These human TPO polypeptides are characterized by a peptide backbone mass of approximately 17,593 to 19,045 amu as determined by mass spectrometry. The polypeptide terminating at Asn (193) has a peptide backbone mass of 18,435 amu. These polypeptides are free of N-linked carbohydrate attachment sites, but contain O-linked carbohydrate attachment sites.

Also provided by the present invention is a TPO polypeptide having an amino acid sequence as shown in SEQ ID NO:4 from Ser (22) to Phe (207).

Protein molecular weights are determined by conventional procedures. See, for example, Laemmli, Nature 227:680-685, 1970; McEwan and Larsen, eds., Mass Spectrometry of Biological Materials, Marcel Decker, Inc., New York, 1990; and Carr et al., Analytical Chem. 63:2802-2824, 1991, which are incorporated herein by reference. Proteins can be deglycosylated enzymatically using conventional methods. In a typical procedure, the protein is desalted by high performance liquid chromatography prior to deglycosylation. N-linked carbohydrate is removed using peptide-N-glycosidase F. Sialic acid residues are removed using sialidase (neuraminidase). O-linked carbohydrate is removed using O-glycosidase (endo-α-N-acetylgalactosaminidase) following removal of terminal neuraminic acid moieties with sialidase. Glycosidases and other enzymes are available from commercial suppliers, such as Oxford Glycosystems (Rosedale, NY) and Boehringer Mannheim, Inc. (Indianapolis, IN).

The present invention provides isolated, biologically active, mammalian TPO polypeptides, including human, mouse, rat, porcine, canine, ovine, bovine and equine TPO polypeptides. Of particular interest are primate TPO polypeptides, in particular human TPO polypeptides. Non-human TPO polypeptides of the present invention are, in general, at least 50% identical in amino acid sequence to corresponding portions of the mouse (SEQ ID NO:2) or human (SEQ ID NO:4) sequences disclosed herein. Percent sequence identity of amino acid sequences is determined by conventional methods. See, for example, Altschul et al., Bull. Math. Bio. 48: 603-616, 1986 and Henikoff and Henikoff, Proc. Natl. Acad. Sci. USA 89:10915-10919, 1992. Briefly, two amino acid sequences

are aligned to optimize the alignment scores using a gap opening penalty of 10, a gap extension penalty of 1, and the "blosum 62" scoring matrix of Henikoff and Henikoff (ibid.) as shown in Table 2 (amino acids are indicated by the standard one-letter codes). The percent identity is then calculated as:

$$\frac{\text{Total number of identical matches}}{\text{[length of the longer sequence plus the number of gaps introduced into the longer sequence in order to align the two sequences]}} \times 100$$

Table 2

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4																			
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	9															
Q	-1	1	0	0	-3	5														
E	-1	0	0	2	-4	2	5													
G	0	-2	0	-1	-3	-2	-2	6												
H	-2	0	1	-1	-3	0	0	-2	8											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

As will be recognized by those skilled in the art, the amino acid sequences shown in SEQ ID NO:2 and SEQ ID NO:4 are merely representative, and allelic variation is expected. Allelic variants are characterized by one or more amino acid substitutions, deletions or additions. In addition, the skilled practitioner will recognize that changes in amino acid sequence can be introduced in TPO molecules through the application of genetic engineering. These engineered changes are preferably of a minor nature, such as conservative amino acid substitutions that do not significantly affect the folding or activity of the protein (see Table 3). See, in general Ford et al., Protein Expression and Purification 2: 95-107, 1991, which is incorporated herein by reference.

Table 3

Conservative amino acid substitutions

Basic:	arginine
	lysine
	histidine
Acidic:	glutamic acid
	aspartic acid
Polar:	glutamine
	asparagine
Hydrophobic:	leucine
	isoleucine
	valine
Aromatic:	phenylalanine
	tryptophan
	tyrosine
Small:	glycine
	alanine
	serine
	threonine
	methionine

For the purpose of engineering variant molecules, essential amino acids in TPO may be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, Science 244, 1081-1085, 1989). In the latter technique, single alanine mutations are introduced at every residue in the molecule, and the resultant mutant molecules are tested for biological activity (e.g. receptor binding, *in vitro* or *in vivo* proliferative activity) to identify amino acid residues that are critical to the activity of the molecule. Sites of ligand-receptor interaction can also be determined by analysis of crystal structure as determined by such techniques as nuclear magnetic resonance, crystallography or photoaffinity labeling. See, for example, de Vos et al., Science 255:306-312, 1992; Smith et al., J. Mol. Biol. 224:899-904, 1992; Wlodaver et al., FEBS Lett. 309:59-64, 1992.

In general, cytokines are predicted to have a four-alpha helix structure, with the first and fourth helices being most important in ligand-receptor interactions and more highly conserved among members of the family. Referring to the human TPO amino acid sequence shown in SEQ ID NO:4, alignment of cytokine sequences suggests that these helices are bounded by amino acid residues 29 and 53, 80 and 99, 108 and 130, and 144 and 168, respectively (boundaries are ± 4 residues). Helix boundaries of the mouse (SEQ ID NO:2) and other non-human TPOs can be determined by alignment with the human sequence. Other important structural aspects of TPO include the cysteine residues at positions 51, 73, 129 and 195 of SEQ ID NO:2 (corresponding to positions 28, 50, 106 and 172 of SEQ ID NO:4).

It is preferred to prepare the TPO polypeptides of the present invention by production in genetically engineered cells or organisms, although isolation from

natural sources of TPO (e.g. blood or other biological fluid) is within the scope of the invention. When producing TPO in a genetically engineered host, a DNA segment encoding at least the polypeptide of interest is expressed in the host, and the polypeptide is recovered. The DNA segment may encode additional sequences, for example additional TPO amino acid residues C-terminal to the carboxyl terminus of a TPO polypeptide or a secretory signal sequence. Within one embodiment of the invention, a DNA segment encoding a full-length TPO molecule is expressed, and the resulting TPO is recovered and fractionated to provide a purified preparation of the truncated polypeptide(s). Within another embodiment, a stop codon is introduced (e.g. by site-specific mutagenesis) into the TPO-encoding DNA segment immediately 3' to the terminal codon of the sequence of interest. Within another embodiment, a DNA segment encoding a TPO polypeptide further encodes one or more non-TPO amino acid residues, such as an amino-terminal methionine, a small amino- or carboxyl-terminal extension that provides an antigenic epitope or other binding domain to facilitate purification of the polypeptide, or a heterologous secretory signal sequence.

In general, a DNA segment encoding TPO is operably linked to a transcription promoter and terminator within an expression vector. The vector will commonly contain one or more selectable markers and one or more origins of replication, although those skilled in the art will recognize that within certain systems selectable markers may be provided on separate vectors, and replication of the exogenous DNA may be provided by integration into the host cell genome. Selection of promoters, terminators, selectable markers, vectors and other elements is a matter of routine design within the level of ordinary skill in the art. Many such elements

are described in the literature and are available through commercial suppliers.

To direct recombinant TPO into the secretory pathway of a host cell, a signal sequence is provided in the expression vector. The signal sequence is joined to the DNA sequence encoding TPO in the correct reading frame. Signal sequences are commonly positioned 5' to the DNA sequence encoding the protein or polypeptide of interest, although certain signal sequences may be positioned elsewhere in the DNA sequence of interest (see, e.g., Welch et al., U.S. Patent No. 5,037,743; Holland et al., U.S. Patent No. 5,143,830). The signal sequence may be one normally associated with TPO, or may be from a gene encoding another secreted protein, such as tissue-type plasminogen activator (t-PA). When using a signal sequence encoding a secretory peptide containing internal proteolytic cleavage sites, it is preferred to eliminate such internal sites through mutagenesis.

Yeast cells, particularly cells of the genus *Saccharomyces*, are a preferred host for use in producing recombinant TPO. Methods for transforming yeast cells with exogenous DNA and producing recombinant proteins therefrom are disclosed by, for example, Kawasaki, U.S. Patent No. 4,599,311; Kawasaki et al., U.S. Patent No. 4,931,373; Brake, U.S. Patent No. 4,870,008; Welch et al., U.S. Patent No. 5,037,743; and Murray et al., U.S. Patent No. 4,845,075, which are incorporated herein by reference. Transformed cells are selected by phenotype determined by the selectable marker, commonly drug resistance or the ability to grow in the absence of a particular nutrient (e.g. leucine). A preferred vector system for use in yeast is the POT1 vector system disclosed by Kawasaki et al. (U.S. Patent No. 4,931,373), which allows transformed cells to be selected by growth in glucose-containing media. A preferred secretory signal sequence for use in yeast is that of the *S. cerevisiae MFα1* gene (Brake, ibid.;

Kurjan et al., U.S. Patent No. 4,546,082). Suitable promoters and terminators for use in yeast include those from glycolytic enzyme genes (see, e.g., Kawasaki, U.S. Patent No. 4,599,311; Kingsman et al., U.S. Patent No. 4,615,974; and Bitter, U.S. Patent No. 4,977,092, which are incorporated herein by reference) and alcohol dehydrogenase genes. See also U.S. Patents Nos. 4,990,446; 5,063,154; 5,139,936 and 4,661,454, which are incorporated herein by reference. Transformation systems for other yeasts, including *Hansenula polymorpha*, *Schizosaccharomyces pombe*, *Kluyveromyces lactis*, *Kluyveromyces fragilis*, *Ustilago maydis*, *Pichia pastoris*, *Pichia guillermondii* and *Candida maltosa* are known in the art. See, for example, Gleeson et al., J. Gen. Microbiol. 132:3459-3465, 1986; Cregg, U.S. Patent No. 4,882,279; and Stroman et al., U.S. Patent No. 4,879,231.

Other fungal cells are also suitable as host cells. For example, *Aspergillus* cells may be utilized according to the methods of McKnight et al., U.S. Patent No. 4,935,349, which is incorporated herein by reference. Methods for transforming *Acremonium chrysogenum* are disclosed by Sumino et al., U.S. Patent No. 5,162,228, which is incorporated herein by reference. Methods for transforming *Neurospora* are disclosed by Lambowitz, U.S. Patent No. 4,486,533, which is incorporated herein by reference.

Cultured mammalian cells are also preferred hosts within the present invention. Methods for introducing exogenous DNA into mammalian host cells include calcium phosphate-mediated transfection (Wigler et al., Cell 14:725, 1978; Corsaro and Pearson, Somatic Cell Genetics 7:603, 1981: Graham and Van der Eb, Virology 52:456, 1973), electroporation (Neumann et al., EMBO J. 1:841-845, 1982), DEAE-dextran mediated transfection (Ausubel et al., eds., Current Protocols in Molecular Biology, John Wiley and Sons, Inc., NY, 1987), and

cationic lipid-mediated transfection (Hawley-Nelson et al., Focus 15:73-79, 1993), which are incorporated herein by reference. The production of recombinant proteins in cultured mammalian cells is disclosed, for example, by Levinson et al., U.S. Patent No. 4,713,339; Hagen et al., U.S. Patent No. 4,784,950; Palmiter et al., U.S. Patent No. 4,579,821; and Ringold, U.S. Patent No. 4,656,134, which are incorporated herein by reference. Preferred cultured mammalian cells include the COS-1 (ATCC No. CRL 1650), COS-7 (ATCC No. CRL 1651), BHK (ATCC No. CRL 1632), BHK 570 (ATCC No. CRL 10314), 293 (ATCC No. CRL 1573; Graham et al., J. Gen. Virol. 36:59-72, 1977) and Chinese hamster ovary (e.g. CHO-K1; ATCC No. CCL 61) cell lines. Additional suitable cell lines are known in the art and available from public depositories such as the American Type Culture Collection, Rockville, Maryland. In general, strong transcription promoters are preferred, such as promoters from SV-40 or cytomegalovirus. See, e.g., U.S. Patent No. 4,956,288. Other suitable promoters include those from metallothionein genes (U.S. Patent Nos. 4,579,821 and 4,601,978, which are incorporated herein by reference) and the adenovirus major late promoter.

Drug selection is generally used to select for cultured mammalian cells into which foreign DNA has been inserted. Such cells are commonly referred to as "transfectants". Cells that have been cultured in the presence of the selective agent and are able to pass the gene of interest to their progeny are referred to as "stable transfectants." A preferred selectable marker is a gene encoding resistance to the antibiotic neomycin. Selection is carried out in the presence of a neomycin-type drug, such as G-418 or the like. Selection systems may also be used to increase the expression level of the gene of interest, a process referred to as "amplification." Amplification is carried out by culturing transfectants in the presence of a low level of

the selective agent and then increasing the amount of selective agent to select for cells that produce high levels of the products of the introduced genes. A preferred amplifiable selectable marker is dihydrofolate reductase, which confers resistance to methotrexate. Other drug resistance genes (e.g. hygromycin resistance, multi-drug resistance, puromycin acetyltransferase) can also be used.

Other higher eukaryotic cells can also be used as hosts, including insect cells, plant cells and avian cells. Transformation of insect cells and production of foreign proteins therein is disclosed by Guarino et al., U.S. Patent No. 5,162,222; Bang et al., U.S. Patent No. 4,775,624; and WIPO publication WO 94/06463, which are incorporated herein by reference. The use of *Agrobacterium rhizogenes* as a vector for expressing genes in plant cells has been reviewed by Sinkar et al., J. Biosci. (Bangalore) 11:47-58, 1987.

Transformed or transfected host cells are cultured according to conventional procedures in a culture medium containing nutrients and other components required for the growth of the chosen host cells. A variety of suitable media, including defined media and complex media, are known in the art and generally include a carbon source, a nitrogen source, essential amino acids, vitamins and minerals. Media may also contain such components as growth factors or serum, as required. The growth medium will generally select for cells containing the exogenously added DNA by, for example, drug selection or deficiency in an essential nutrient which is complemented by the selectable marker carried on the expression vector or co-transfected into the host cell.

Transgenic animal technology may also be employed to produce TPO. It is preferred to produce the protein within the mammary glands of a host female mammal. Expression in the mammary gland and subsequent secretion

of the protein of interest into the milk overcomes many difficulties encountered in isolating proteins from other sources. Milk is readily collected, available in large quantities, and well characterized biochemically. Furthermore, the major milk proteins are present in milk at high concentrations (from about 1 to 15 g/l).

From a commercial point of view, it is clearly preferable to use as the host a species that has a large milk yield. While smaller animals such as mice and rats can be used (and are preferred at the proof-of-concept stage), it is preferred to use livestock mammals including, but not limited to, pigs, goats, sheep and cattle. Sheep are particularly preferred due to such factors as the previous history of transgenesis in this species, milk yield, cost and the ready availability of equipment for collecting sheep milk. See WIPO Publication WO 88/00239 for a comparison of factors influencing the choice of host species. It is generally desirable to select a breed of host animal that has been bred for dairy use, such as East Friesland sheep, or to introduce dairy stock by breeding of the transgenic line at a later date. In any event, animals of known, good health status should be used.

To obtain expression in the mammary gland, a transcription promoter from a milk protein gene is used. Milk protein genes include those genes encoding caseins (see U.S. Patent No. 5,304,489, incorporated herein by reference), beta-lactoglobulin, α-lactalbumin, and whey acidic protein. The beta-lactoglobulin (BLG) promoter is preferred. In the case of the ovine beta-lactoglobulin gene, a region of at least the proximal 406 bp of 5' flanking sequence of the gene will generally be used, although larger portions of the 5' flanking sequence, up to about 5 kbp, are preferred, such as a -4.25 kbp DNA segment encompassing the 5' flanking promoter and non-coding portion of the beta-lactoglobulin gene. See

Whitelaw et al., Biochem J. 286: 31-39, 1992. Similar fragments of promoter DNA from other species are also suitable.

General procedures for producing transgenic animals are known in the art. See, for example, Hogan et al., Manipulating the Mouse Embryo: A Laboratory Manual, Cold Spring Harbor Laboratory, 1986; Simons et al., Bio/Technology 6: 179-183, 1988; Wall et al., Biol. Reprod. 32: 645-651, 1985; Buhler et al., Bio/Technology 8: 140-143, 1990; Ebert et al., Bio/Technology 9: 835-838, 1991; Krimpenfort et al., Bio/Technology 9: 844-847, 1991; Wall et al., J. Cell. Biochem. 49: 113-120, 1992; U.S. Patents Nos. 4,873,191 and 4,873,316; WIPO publications WO 88/00239, WO 90/05188, WO 92/11757; and GB 87/00458, which are incorporated herein by reference. Techniques for introducing foreign DNA sequences into mammals and their germ cells were originally developed in the mouse. See, e.g., Gordon et al., Proc. Natl. Acad. Sci. USA 77: 7380-7384, 1980; Gordon and Ruddle, Science 214: 1244-1246, 1981; Palmiter and Brinster, Cell 41: 343-345, 1985; Brinster et al., Proc. Natl. Acad. Sci. USA 82: 4438-4442, 1985; and Hogan et al. (ibid.). These techniques were subsequently adapted for use with larger animals, including livestock species (see e.g., WIPO publications WO 88/00239, WO 90/05188, and WO 92/11757; and Simons et al., Bio/Technology 6: 179-183, 1988). To summarize, in the most efficient route used to date in the generation of transgenic mice or livestock, several hundred linear molecules of the DNA of interest are injected into one of the pro-nuclei of a fertilized egg according to techniques which have become standard in the art. Injection of DNA into the cytoplasm of a zygote can also be employed.

Production in transgenic plants may also be employed. Expression may be generalized or directed to a particular organ, such as a tuber. See, Hiatt, Nature 344:469-479, 1990; Edelbaum et al., J. Interferon Res.

12:449-453, 1992; Sijmons et al., Bio/Technology 8:217-221, 1990; and European Patent Office Publication EP 255,378.

Within one embodiment of the present invention, low molecular weight TPO is selectively recovered from cell culture media using a combination of ultrafiltration, affinity chromatography and ion exchange chromatography. Additional purification procedures may be employed, such as hydrophobic interaction chromatography. In a typical purification, conditioned medium from cells expressing recombinant TPO, which contains a mixture of TPO species from about 17.5 kD to about 70 kD, is concentrated. Suitable means of concentration include binding to a dye resin and ultrafiltration, the latter being preferred. The degree of concentration will be determined by such factors as initial volume and downstream processing capacity. In general, approximately 20-fold concentration is preferred, with a practical upper limit of about 50-fold in most cases.

The concentrated medium is then combined with an immobilized MPL receptor polypeptide comprising at least the ligand-binding domain of the receptor. The ligand binding domain of the mouse MPL receptor is contained within the extracellular portion of the protein (residues 27 to 480 of SEQ ID NO: 9), with residues 293-297, 358-361, and 398-419 believed to be of particular importance for ligand binding. In a typical procedure, concentrated medium is adjusted to 0.5 M NaCl and a slightly alkaline pH (preferably about pH 8.0), and the column is equilibrated with a similar buffer. The medium is then applied to the column, and bound TPO polypeptide is eluted with 3 M KSCN, pH 8.0, or 3 M KSCN, 1.0 M NaCl in 0.1 M borate, pH 10.0.

Material eluted from the immobilized receptor is then dialyzed to remove salt, such as in 3-4 changes of 20 mM Tris pH 8.5. The dialyzed sample is then fractionated

on a strong anion exhange medium. Suitable anion exchange media for use in this procedure include polymeric (e.g., agarose, dextran or polystyrene) beads derivatized with quaternary amino groups, used in a fast pressure liquid chromatography format. A preferred anion exchanger is Mono-Q Sepharose (available from Pharmacia Biotech, Piscataway, NJ). Bound TPO is eluted from the anion exchange medium by applying a salt gradient. When using Mono-Q Sepharose, the 18.5 kD species typically elutes at a salt concentration of approximately 0.15 M NaCl.

Within another embodiment of the present invention, a truncated TPO DNA is expressed in an engineered host cell. The resulting TPO polypeptide is secreted into the culture media, the media and cells are separated, and the TPO polypeptide is selectively recovered from the media. The TPO polypeptide can be recovered by a combination of affinity purification and other techniques as disclosed above.

The TPO of the present invention can be used therapeutically wherever it is desirable to increase proliferation of cells in the bone marrow, such as in the treatment of cytopenia, such as that induced by aplastic anemia, myelodisplastic syndromes, chemotherapy or congenital cytopenias. The proteins are also useful for increasing platelet production, such as in the treatment of thrombocytopenia. Thrombocytopenia is associated with a diverse group of diseases and clinical situations that may act alone or in concert to produce the condition. Lowered platelet counts can result from, for example, defects in platelet production, abnormal platelet distribution, dilutional losses due to massive transfusions, or abnormal destruction of platelets. For example, chemotherapeutic drugs used in cancer therapy may suppress development of platelet progenitor cells in the bone marrow, and the resulting thrombocytopenia limits the chemotherapy and may necessitate transfusions. In

addition, certain malignancies can impair platelet production and platelet distribution. Radiation therapy used to kill malignant cells also kills platelet progenitor cells. Thrombocytopenia may also arise from various platelet autoimmune disorders induced by drugs, neonatal alloimmunity or platelet transfusion alloimmunity. The proteins of the present invention can reduce or eliminate the need for transfusions, thereby reducing the incidence of platelet alloimmunity. Abnormal destruction of platelets can result from: (1) increased platelet consumption in vascular grafts or traumatized tissue; or (2) immune mechanisms associated with, for example, drug-induced thrombocytopenia, idiopathic thrombocytopenic purpura (ITP), autoimmune diseases, hematologic disorders such as leukemia and lymphoma, or metastatic cancers involving bone marrow. Other indications for the proteins of the present invention include aplastic anemia and drug-induced marrow suppression resulting from, for example, chemotherapy or treatment of HIV infection with AZT.

Thrombocytopenia is manifested as increased bleeding, such as mucosal bleedings from the nasal-oral area or the gastrointestinal tract, as well as oozing from wounds, ulcers or injection sites.

For pharmaceutical use, TPO is formulated for parenteral, particularly intravenous or subcutaneous, delivery according to conventional methods. Intravenous administration will be by bolus injection or infusion over a typical period of one to several hours. In general, pharmaceutical formulations will include TPO in combination with a pharmaceutically acceptable vehicle, such as saline, buffered saline, 5% dextrose in water or the like. Formulations may further include one or more excipients, preservatives, solubilizers, buffering agents, albumin to prevent protein loss on vial surfaces, etc. In addition, TPO can be combined with other cytokines,

particularly early-acting cytokines such as stem cell factor, IL-3, IL-6, IL-11 or GM-CSF. When utilizing such a combination therapy, the cytokines may be combined in a single formulation or may be administered in separate formulations. Methods of formulation are well known in the art and are disclosed, for example, in Remington's Pharmaceutical Sciences, Gennaro, ed., Mack Publishing Co., Easton PA, 1990, which is incorporated herein by reference. Therapeutic doses of the TPO of the present invention will generally be in the range of 0.1 to 100 μg/kg of patient weight per day, preferably 0.5-50 μg/kg per day, with the exact dose determined by the clinician according to accepted standards, taking into account the nature and severity of the condition to be treated, patient traits, etc. In certain cases, such as when treating patients showing increased sensitivity or requiring prolonged treatment, doses in the range of 0.1-20 μg/kg per day will be indicated. Determination of dose is within the level of ordinary skill in the art. TPO will commonly be administered over a period of up to 28 days following chemotherapy or bone-marrow transplant or until a platelet count of >20,000/mm^3, preferably >50,000/mm^3, is achieved. More commonly, TPO will be administered over one week or less, often over a period of one to three days. In general, a therapeutically effective amount of TPO is an amount sufficient to produce a clinically significant increase in the proliferation and/or differentiation of lymphoid or myeloid progenitor cells, which will be manifested as an increase in circulating levels of mature cells (e.g. platelets or neutrophils). Treatment of platelet disorders will thus be continued until a platelet count of at least 20,000/mm^3, preferably 50,000/mm^3, is reached. TPO can also be administered in combination with other cytokines such as IL-3, -6 and -11; stem cell factor; erythropoietin; G-CSF and GM-CSF. Within regimens of

combination therapy, daily doses of other cytokines will in general be: EPO, ≤ 150 U/kg; GM-CSF, 5-15 μg/kg; IL-3, 1-5 μg/kg; and G-CSF, 1-25 μg/kg. Combination therapy with EPO, for example, is indicated in anemic patients with low EPO levels, wherein EPO is administered in an amount sufficient to increase erythropoiesis. Increased erythropoiesis is manifested as a subsequent increase in hematocrit.

TPO is also a valuable tool for the *in vitro* study of the differentiation and development of hematopoietic cells, such as for elucidating the mechanisms of cell differentiation and for determining the lineages of mature cells, and may also find utility as a proliferative agent in cell culture.

TPO can also be used *ex vivo*, such as in autologous marrow culture. Briefly, bone marrow is removed from a patient prior to chemotherapy and treated with TPO, optionally in combination with one or more other cytokines. The treated marrow is then returned to the patient after chemotherapy to speed the recovery of the marrow. In addition, TPO can be used for the *ex vivo* expansion of marrow or peripheral blood progenitor (PBPC) cells. Prior to chemotherapy treatment, marrow can be stimulated with stem cell factor (SCF) or G-CSF to release early progenitor cells into peripheral circulation. These progenitors can be collected and concentrated from peripheral blood and then treated in culture with TPO, optionally in combination with one or more other cytokines, including but not limited to SCF, G-CSF, IL-3, GM-CSF, IL-6 or IL-11, to differentiate and proliferate into high-density megakaryocyte cultures, which can then be returned to the patient following high-dose chemotherapy.

The invention is further illustrated by the following non-limiting examples.

Examples

Example 1

Plasmid pZGmpl-1081 (deposited under the terms of the Budapest Treaty with American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD on Fegruary 14, 1994 as an *E. coli* DH5α transformant and assigned accession number 69566) was digested with Eco RI and Not I, and the TPO DNA segment was recovered. This DNA was inserted into Eco RI-digested, alkaline phosphatase-treated plasmid Zem229R with a Not I/Eco RI linker. (Zem229R was deposited under the terms of the Budapest Treaty with American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD on September 23, 1993 as an *E. coli* HB101 transformant and assigned accession number 69447.) The resulting plasmid, designated mpl.229R, was transfected into BHK 570 cells (ATCC CRL 10314). The transfectants were grown in 10-layer cell factories (Nunc, Inc.; obtained from VWR Scientific, Seattle, WA) in serum-free medium and selected 1 μM methotrexate. Sixteen liters of conditioned culture medium was collected.

TPO was purified from the conditioned medium by affinity chromatography on immobilized MPL receptor. Eighty mg of purified mouse MPL receptor extracellular domain (see Example 7) was immobilized on 8 ml of Affi-Prep 10 acrylic polymer support (Bio-Rad Laboratories, Inc.) using procedures specified by the manufacturer. The receptor-support matrix was packed into an 8 ml column.

The conditioned cell culture media, which contained a prominent TPO band of approximately M_r=18.5 kD, was concentrated 29-fold on a 10 Kd cut off hollow fiber membrane (A/G Technology Corp., Needham, MA). The concentrated media was cycled over the affinity column for several hours or overnight in the cold. After washing the column to baseline absorbance (at 280 nm) with 20 mM Tris pH 8.5, the column was eluted with 3 M KSCN (Fluka

Chemical Corp., Ronkonkoma, NY). Eluate fractions were pooled and dialyzed against 20 mM Tris pH 8.5.

The dialyzed protein was applied to a Mono-Q Sepharose column (Pharmacia Biotech, Piscataway, NJ) at pH 8.5. Bound TPO was eluted from the column with a NaCl gradient (0 to 0.5 M). Peak fractions were identified by monitoring absorbance at 220 nm and "western" blotting.

Units of TPO activity were determined by assaying mitogenic activity on a TPO-dependent cell line. A BHK 570 cell line transfected with the mouse TPO expression vector pZGmpl-1081 was grown in serum-free medium. Conditioned culture medium was collected, and an asymptotic mitogenic activity curve was generated using this standard solution. The target cells were BaF3/MPLR1.1 cells (IL-3-dependent cells expressing a stably transfected Type I mouse MPL receptor; deposited September 28, 1994 under the terms of the Budapest Treaty with American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD and assigned accession number CRL 11723). The point of 1/2 maximal activity (average of 16 curves) was assigned the value of 50 U/ml. The original standard solution was calculated to contain 26,600 U/ml mouse TPO.

For test samples, a culture supernatant or purified protein preparation was diluted in RPMI 1640 medium supplemented with 57 μM 2-mercaptoethanol, 2 mM L-glutamine, 1 mM sodium pyruvate, PSN antibiotic mixture, 10 mM HEPES and 10% heat inactivated fetal bovine serum, generally using 8-24 dilutions. Briefly, 100 μl of diluted test sample or standard sample and 100 μl BaF3 cells (final cell number added about 10,000 cells/well) were combined in wells of a 96 well plate. Internal standards included eight 2-fold dilutions of 100 U/ml mouse TPO for mouse TPO assays, or eight 2-fold dilutions of 150 U/ml mouse TPO for human TPO assays. To each well was added 2 μl ^{3}H-

thymidine (1 μCi/μl; Amersham), and the plates were incubated overnight at 37°C.

The contents of each well of each plate were transferred to a filter/plate using a Packard apparatus. The filters were washed 8 times with water, and the filters were dried and counted. Units of TPO activity in each sample well were determined by comparison to the standard curve.

The activity of the protein was determined to be 14×10^6 U/ml. Protein concentration was 30 μg/ml.

Example 2

Thrombopoietin polypeptide from the preparation disclosed in Example 1 ran at approximately 20 kDa on a non-reducing SDS PAGE gel (less than the 35,594.9 amu predicted on the basis of the amino acid sequence) and was active *in vitro*. Edman degradation indicated that the N-terminus was intact. These data demonstrate that the TPO had been proteolytically processed at the C-terminus.

To determine the C-terminal processing site of the ca. Mr=18.5 kD mouse TPO, the polypeptide was further analyzed by enzymatic deglycosylation and mass spectrometry. A sequential deglycosylation experiment was set up to determine if N or/and O-glycosylation were present. Results of this experiment would tell if the processing site was before or after the first N-linked glycosylation site. In addition, deglycosylation of all N and O-linked sugars was necessary to carry out mass spectrometry for determination of the C-terminus.

A sample of the polypeptide was desalted by reverse phase HPLC. 222.6 μg of sample was desalted on a 2.1 mm X 150 mm PLRP-S 4000 column (Polymer Laboratories, Inc., Amherst, MA) in two separate runs of 111 μg each using conditions shown in Table 4.

Table 4

HPLC: Michrom Bioresource Ultrafast Michroprotein Analyzer.

Solvent A : 2% CH_3CN / 0.01% TFA/ H_2O.

Solvent B: 10% H_2O / 90% CH_3CN / 0.09% TFA.

Gradient:

Time(minutes)	% Solvent B
0	10
20	98
22	98
25	10

Wavelength: 215 nm; Flow 0.5 ml/minute.

The eluates were collected manually. The final volume of the eluate was 1.57 ml. Subvials of 15 X 100 µl (approximately 14.1 µg per 100 µl vial) and 1 X 70 µl were concentrated in a vacuum concentrator (Savant Speed Vac) and stored at -20°C.

Sequential digestion of the sample was performed with peptide-N-glycosidase F (PNGase F), sialidase, and O-glycosidase to determine if N- or/and O-linked glycosylation were present. Deglycosylation was carried out in three steps.

To remove N-linked sugars, two aliquots of 28.2 µg of desalted polypeptide were each reconstituted in 100 µl of PNGase F reaction buffer (20 mM sodium phosphate pH 7.5, 50 mM EDTA, and 0.02% sodium azide). Five µl of PNGase F was added to one sample, and the remaining sample was used as a control for the reaction. Both samples were incubated at 37°C for 18 hours. PNGase F (purchased from Oxford Glycosystems, Rosedale, NY at a concentration of 150-200 U/ml) cleaves the beta aspartyl-glycosylamine bond between asparagine and the innermost N-acetylglucosamine of the glycan. As a positive control for this reaction, 1 ug of a mouse TPO known to contain N-linked sugars and migrate at approximately 70 kDa on a SDS PAGE reducing gel

was used. After 18 hours, 2 X 5 µl of the PNGase F digested material was removed and stored at -20°C for gel analysis.

5 µl of sialidase (purchased from Oxford Glycosytems at a concentration of 5 mU per µl) was added to the remainder of the PNGase F-digested material, and the mixture and incubated for two hours at 37°C. After the incubation was complete, two 5 µl aliquots of the digest were stored at -20°C for gel analysis. Sialidase is an exoglycosidase that specifically cleaves the terminal neuraminic acid. This reaction is required prior to cleaving O-linked sugars in mammalian proteins.

5 µl (1.5 mU) of O-glycosidase (Oxford Glycosystems; 300 mU/ml) was added to the remainder of the digested sample and then incubated for 18 hours at 37°C. O-glycosidase is specific for liberating Galβ$_1$-3GalNAc from serine or threonine.

The sequential digestion was then analyzed by silver stain and western blot analysis. Samples were electrophoresed on 4-20% reducing SDS PAGE gels, 1.0 mm in thickness (Novex, San Diego, CA). Gels were stained with Novex Silver Express stain. For western blot analysis, rat anti-mouse TPO antisera (1:1000 dilution) was used. The western blot and silver stain results showed that there were no N-linked glycosylation sites on the mouse TPO polypeptide, indicating that the cleavage site of this form must be N-terminal to asparagine-220 of SEQ ID NO:2. The positive control showed a shift from 70 kDa to 45 kDa. After digestion with O-glycanase, there was a shift in molecular weight of the truncated polypeptide, indicating the presence of O-glycosylation.

Mass spectral analysis was performed on 28 µg of the polypeptide in three samples: with O-linked sugars present; after removal of sialic acid; and after sequential deglycosylation. The samples were all desalted prior to analysis and reconstituted in 100 µl of a 1:1

ratio of 0.1% formic acid/H_2O and 0.08% formic acid in ethanol/propanol (5:2). Electrospray mass spectra were recorded on a Sciex (Thornhill, Ontario) API III triple quadrupole mass spectrometer fitted with an articulated ionspray plenum and an atmospheric pressure ionization source. The mass spectrometer was tuned and calibrated using a mixture of polypropylene glycols (PPG) 425, 1,000, and 2,000 (3.3×10^{-5} M, 1×10^{-4} M, and 2×10^{-4} M, respectively), in 50/50/0.1 H_2O/methanol/formic acid (v/v/v), 1 mM NH_4OAc. Normal scan ESMS were recorded at instrument conditions sufficient to resolve isomers of the PPG/NH_4^+ doubly charged ion at m/z 520 (85% valley definition). Each sample was infused at 5 µl per minute with up to 81 scans averaged. The mass spectrometer was scanned over a range of m/z 1200-2400, with a dwell time of 2 mSec and a step size of 0.1 amu. The orvice potential was set at 150 V.

Reconstructed ion chromatogram mass spectra of the fully deglycosylated sample showed a heterogeneous C-terminus, starting at serine residue 208 and ending at asparagine residue 216. The masses found take into consideration that 4 amu are subtracted from the predicted mass due to disulfide bond formation and 1 amu is subtracted from the final MH^+ mass shown in the MacBiospec sequence analysis.

The information obtained from mass spectral analysis revealed that each of the O-linked carbohydrate chains on the sample had one of the following compositions, with the addition of up to six sialic acids on the N-terminal sugar of each chain: 5 HexNAc, 5 Hexose; 4 HexNAc, 4 Hexose; 3 HexNAc, 3 Hexose.

Example 3

Recombinant TPO, prepared as disclosed in Example 1 (TPO 711), was tested for thrombopoietic activity in mice.

Female Balb/c mice (obtained from Jackson Laboratories, Bar Harbor, ME) were divided into six groups (five treatment groups and one control group) of four mice each. On day -7, 50 µl of blood was collected by retro-orbital bleed in a Microtainer tube (Becton Dickinson, Rutherford, NJ), and complete blood counts were performed.

Recombinant TPO was administered by intraperitoneal injection on days 1 through 10. Group I received vehicle (prepared by combining 2 ml 1 M Tris pH 8.0 stock, 98 ml USP water for injection, 250 mg rabbit serum albumin and 0.93 g NaCl) only. Group II received 20 kU/day recombinant TPO from cells expressing a full-length mouse TPO cDNA (designated TPO 525). This primary TPO species in this preparation had apparent molecular weights of approximately 30 and 70 kD. Groups III-VI received 5, 10, 20 or 50 kU/day, respectively, of recombinant TPO 711. All mice were bled and blood counts were performed on days 3, 7, 10 and 14. Body weight was monitored on days -7 or -5, 3, 7, 10 and 14. On day 14, mice were anesthetized by injection with a ketamine/Rompun mixture or by inhalation of metafane. Blood was collected, and the animals were sacrificed in a CO_2 chamber or by cervical dislocation.

Results of the study are presented in Table 5. Data for days 3, 7, 10 and 14 are shown as percent of basal (day 0) platelet counts. TPO 711 at 50 kU/day increased platelet levels 2.7 fold on day 10, similar to the peak response seen with TPO 525 on day 7. These results indicate that the truncated TPO 711 is at least as active as TPO 525. A dose-response relationship was seen for TPO 711 on days 10 and 14.

Table 5

Group	Platelet Count, % of Basal (mean ± SD) Day 3	Day 5	Day 10	Day 14
I	97±14	113±15	108±16	117±16
II	100±5	290±59	230±23	249±83
III	93±9	166±23	126±13	129±25
IV	95±7	147±24	173±21	164±22
V	90±22	164±27	176±42	194±31
VI	141±38	173±50	271±46	214±43

Example 4

A second series of experiments was conducted to compare the effects of TPO 711 administered by once daily injection and twice daily injection.

Three groups of female Balb/c mice (Jackson Laboratories) were used. Group I (control; n=3) received twice daily injections of vehicle (prepared by combining 2 ml 1 M Tris pH 8.0 stock, 98 ml USP water for injection, 250 mg rabbit serum albumin and 0.93 g NaCl) alone. Group II (n=4) received daily intraperitoneal injections of 50 kU TPO. Group III received twice daily i.p. injections of 25 kU TPO (50 kU/day).

On day -7, 50 µl of blood was collected from each animal by orbital sinus bleeding, and blood counts were determined. On day 0, injections were begun. Injections were continued through day 10. The animals were bled again on days 3 and 7 for blood counts. On day 10 the animals were sacrificed in a CO_2 chamber, and platelet counts were determined.

Results of this study (shown in Fig. 1) indicated that treatment with 25 kU TPO 711 twice daily increases platelet counts by more than 300% by day 7. Twice-daily administration produced higher platelet counts than the same dose given in a single injection.

Example 5

Experiments were conducted to compare the effects of TPO 711 administered twice daily vs. once daily and to compare intraperitoneal and subcutaneous routes of injection. In addition, a dose-response curve was established.

Female Balb/c mice weighing about eighteen grams each (Jackson Laboratories, Bar Harbor, ME) were divided into seven groups as shown in Table 6. Groups I-V received the indicated dose (determined as disclosed in Example 1) administered as two equal intraperitoneal injections per day. Group VI received a single, daily i.p. injection. Group VII received a single, daily subcutaneous injection.

Table 6

Group	Treatment	n
I	vehicle control	4
II	20 kU TPO/day	4
III	50 kU TPO/day	4
IV	100 kU TPO/day	4
V	200 kU TPO/day	4
VI	100 kU TPO/day	4
VII	100 kU TPO/day	4

On day -7, all animals were bled, and complete blood counts (CBC), differential blood counts, and reticulocyte counts were determined. Treatment began on day 0 and continued through day 10. On day 3, blood was collected from each animal for determination of CBC, differentials and reticulocytes. On day 7, blood was collected from each animal, CBC and differentials were determined, and the animals were weighed. On day 10, each animal was bled, CBC were determined, body weight was measured, and the animals were sacrificed with CO_2.

As shown in Fig. 2, treatment with TPO 711 caused a dose-dependent increase in platelet counts over the 11 day duration of the study.

As shown in Table 7, twice daily administration resulted in a higher platelet response than did once daily administration. Subcutaneous injection once a day resulted in a higher platelet response than did i.p. injection.

Table 7

Group	Platelet Count (Day 10) % of Basal (mean +/- SD)
I	100.7 ± 22.6
II	206.9 ± 32.3
III	292.6 ± 70.8
IV	301.2 ± 73.7
V	351.3 ± 36.9
VI	123.4 ± 27.2
VII	193.3 ± 14.6

Example 6

A vector was constructed for expression of a TPO polypeptide ending at amino acid residue 193 of SEQ ID NO:4. The human TPO DNA sequence was mutagenized by PCR to introduce a stop codon and an EcoRI site following the codon for amino acid 193. Ten ng of template DNA was combined with 5 µl of 2 mM dNTPs, 5 µl 10x Taq buffer (Boehringer Mannheim, Indianapolis, IN), 0.2 µl Taq DNA polymerase (Boehringer Mannheim), 40 pmole of each primer ZC8045 (SEQ ID NO:5) and ZC7878 (SEQ ID NO:6), and H_2O to 50 µl. The mixture was incubated for 30 cycles of 95°C, 1 minute; 50°C, 2 minutes; and 72°C, 1 minute, with a final ten minute incubation at 72°C. DNA was isolated from the reaction mixture and digested with PstI and Eco RI, and a 204 bp fragment encoding amino acid residues 127-193 of

SEQ ID NO:4 was recovered by electrophoresis and centrifugal extraction from a gel slice.

To construct the expression vector, the isolated PCR product was ligated with an EcoRI-PstI fragment encoding a modified human t-PA leader and amino acid residues 22-126 of SEQ ID NO:4, and Zem229R (deposited under the terms of the Budapest Treaty with American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD on September 28, 1993 as an *E. coli* HB101 transformant and assigned Accession Number 69447) that had been digested with EcoRI and treated with alkaline phosphatase. The ligated DNA was used to transform competent *E. coli* DH10b™ cells (GIBCO BRL, Gaithersburg, MD). The plasmid was designated TPO202.229R.

BHK cells were transfected with TPO202.229R using a 3:1 liposome formulation of 2,3-dioleyloxy-N-[2(sperminecarboxyamido)ethyl]-N,N-dimethyl-1-propanaminiumtrifluoroacetate and dioleoly-phosphatidylethanolamine in water (Lipofectamine™ reagent, GIBCO). Transfectants were selected in 500 nM methotrexate (MTX). Pooled cells produced 13,110 U/ml TPO. After amplification in 5 μM MTX, pooled cells produced 20,850 U/ml/day TPO.

The 646 bp EcoRI insert was removed from TPO202.229R and ligated to the vector pDX (disclosed in U.S. Patent No. 4,959,318) that had been linearized by digestion with EcoRI and treated with alkaline phosphatase. The resulting vector, designated TPO202.pDX, was cotransfected into BHK 570 cells with Zem229R. Cells amplified in 500 nM MTX produced 17,000 U/ml/day TPO.

Example 7

Spleens from C57BL/KsJ-db/db mice were removed and immediately placed in liquid nitrogen. Total RNA was prepared from spleen tissue using guanidine isothiocyanate (Chirgwin et al., Biochemistry 18: 52-94, 1979) followed

by a CsCl centrifugation step. Spleen poly(A)+ RNA was isolated using oligo d(T) cellulose chromatography (Aviv and Leder, Proc. Natl. Acad. Sci. U.S.A. 69: 1408-1412, 1972).

Seven and a half μl of poly d(T)-selected poly(A)$^+$ mouse spleen RNA at a concentration of 1.7 μg/μl was mixed with 3 μl of 20 pmole/μl first strand primer ZC6091 (SEQ ID NO:7) containing a *Not* I restriction site. The mixture was heated at 65° C for 4 minutes and cooled by chilling on ice. First strand cDNA synthesis was initiated by the addition of 8 μl of 250 mM Tris-HCl, pH 8.3, 375 mM KCl, 15 mM $MgCl_2$ (5x SUPERSCRIPT™ buffer; GIBCO BRL), 4 μl of 100 mM dithiothreitol and 3 μl of a deoxynucleotide triphosphate solution containing 10 mM each of dATP, dGTP, dTTP and 5-methyl-dCTP (Pharmacia LKB Biotechnology, Inc., Piscataway, NJ) to the RNA-primer mixture. The reaction mixture was incubated at 45° C for 4 minutes followed by the addition of 10 μl of 200 U/μl RNase H$^-$ reverse transcriptase (GIBCO BRL, Gaithersburg, MD). The efficiency of the first strand synthesis was analyzed in a parallel reaction by the addition of 10 μCi of ^{32}P-αdCTP to a 10 μl aliquot of the reaction mixture to label the reaction for analysis. The reactions were incubated at 45° C for 1 hour followed by an incubation at 50° C for 15 minutes. Unincorporated ^{32}P-αdCTP in the labeled reaction was removed by chromatography on a 400 pore size gel filtration column (CHROMA SPIN + TE-400™; Clontech Laboratories Inc., Palo Alto, CA). Unincorporated nucleotides in the unlabeled first strand reaction were removed by twice precipitating the cDNA in the presence of 8 μg of glycogen carrier, 2.5 M ammonium acetate and 2.5 volume ethanol. The unlabeled cDNA was resuspended in 50 μl water for use in second strand synthesis. The length of the labeled first strand cDNA was determined by agarose gel electrophoresis.

Second strand synthesis was performed on first strand cDNA under conditions that promoted first strand priming of second strand synthesis resulting in DNA hairpin formation. The reaction mixture was assembled at room temperature and consisted of 50 µl of the unlabeled first strand cDNA, 16.5 µl water, 20 µl of 5x polymerase I buffer (100 mM Tris: HCl, pH 7.4, 500 mM KCl, 25 mM $MgCl_2$, 50 mM $(NH_4)_2SO_4$), 1 µl of 100 mM dithiothreitol, 2 µl of a solution containing 10 mM of each deoxynucleotide triphosphate, 3 µl of 5 mM β-NAD, 15 µl of 3 U/µl *E. coli* DNA ligase (New England Biolabs Inc., Beverly, MA) and 5 µl of 10 U/µl *E. coli* DNA polymerase I (Amersham Corp., Arlington Heights, IL). The reaction was incubated at room temperature for 5 minutes followed by the addition of 1.5 µl of 2 U/µl RNase H (GIBCO BRL). A parallel reaction in which a 10 µl aliquot of the second strand synthesis mixture was labeled by the addition of 10 µCi ^{32}P-αdCTP was used to monitor the efficiency of second strand synthesis. The reactions were incubated at 15° C for two hours followed by a 15 minute incubation at room temperature. Unincorporated ^{32}P-αdCTP in the labeled reaction was removed by chromatography through a 400 pore size gel filtration column (Clontech Laboratories, Inc.) before analysis by agarose gel electrophoresis. The unlabeled reaction was terminated by two extractions with phenol/chloroform and a chloroform extraction followed by ethanol precipitation in the presence of 2.5 M ammonium acetate.

The single-stranded DNA of the hairpin structure was cleaved using mung bean nuclease. The reaction mixture contained 100 µl of second strand cDNA, 20 µl of 10x mung bean nuclease buffer (Stratagene Cloning Systems, La Jolla, CA), 16 µl of 100 mM dithiothreitol, 51.5 µl of water and 12.5 µl of a 1:10 dilution of mung bean nuclease (Promega Corp.; final concentration 10.5 U/µl) in mung bean nuclease dilution buffer. The reaction was incubated at

37° C for 15 minutes. The reaction was terminated by the addition of 20 µl of 1 M Tris-HCl, pH 8.0 followed by sequential phenol/chloroform and chloroform extractions as described above. Following the extractions, the DNA was precipitated in ethanol and resuspended in water.

The resuspended cDNA was blunt-ended with T4 DNA polymerase. The cDNA, which was resuspended in 190 µl of water, was mixed with 50 µl 5x T4 DNA polymerase buffer (250 mM Tris-HCl, pH 8.0, 250 mM KCl, 25 mM $MgCl_2$), 3 µl 0.1 M dithiothreitol, 3 µl of a solution containing 10 mM of each deoxynucleotide triphosphate and 4 µl of 1 U/µl T4 DNA polymerase (Boehringer Mannheim Corp., Indianapolis, IN). After an incubation of 1 hour at 10° C, the reaction was terminated by the addition of 10 µl of 0.5 M EDTA followed by serial phenol/chloroform and chloroform extractions as described above. The DNA was chromatographed through a 400 pore size gel filtration column (Clontech Laboratories Inc.) to remove trace levels of protein and to remove short cDNAs less than ~400 bp in length. The DNA was ethanol precipitated in the presence of 12 µg glycogen carrier and 2.5 M ammonium acetate and was resuspended in 10 µl of water. Based on the incorporation of ^{32}P-αdCTP, the yield of cDNA was estimated to be ~2 µg from a starting mRNA template of 12.5 µg.

Eco RI adapters were ligated onto the 5' ends of the cDNA to enable cloning into a lambda phage vector. A 10 µl aliquot of cDNA (~2µg) and 10 µl of 65 pmole/µl of *Eco* RI adapter (Pharmacia LKB Biotechnology Inc.) were mixed with 2.5 µl 10x ligase buffer (Promega Corp.), 1 µl of 10 mM ATP and 2 µl of 15 U/µl T4 DNA ligase (Promega Corp.). The reaction was incubated overnight (~18 hours) at a temperature gradient of 0° C to 18° C. The reaction was further incubated overnight at 12° C. The reaction was terminated by the addition of 75 µl of water and 10 µl of 3 M Na acetate, followed by incubation at 65° C for 30 minutes. After incubation, the cDNA was extracted with

phenol/chloroform and chloroform as described above and precipitated in the presence of 2.5 M ammonium acetate and 1.2 volume of isopropanol. Following centrifugation, the cDNA pellet was washed with 70% ethanol, air dried and resuspended in 89 µl water.

To facilitate the directional cloning of the cDNA into a lambda phage vector, the cDNA was digested with *Not* I, resulting in a cDNA having 5' *Eco* RI and 3' *Not* I cohesive ends. The *Not* I restriction site at the 3' end of the cDNA had been previously introduced through primer ZG6091 (SEQ ID NO:7). Restriction enzyme digestion was carried out in a reaction mixture containing 89 µl of cDNA described above, 10 µl of 6 mM Tris-HCl, 6 mM $MgCl_2$, 150 mM NaCl, 1 mM DTT (10x D buffer; Promega Corp., Madison, WI) and 1 µl of 12 U/µl *Not* I (Promega Corp.). Digestion was carried out at 37° C for 1 hour. The reaction was terminated by serial phenol/chloroform and chloroform extractions. The cDNA was ethanol precipitated, washed with 70% ethanol, air dried and resuspended in 20 µl of 1x gel loading buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 5% glycerol and 0.125% bromphenol blue).

The resuspended cDNA was heated to 65°C for 5 minutes, cooled on ice and electrophoresed on a 0.8% low melt agarose gel (SEA PLAQUE GTG™ low melt agarose; FMC Corp.). Unincorporated adapters and cDNA below 1.6 kb in length were excised from the gel. The electrodes were reversed, and the cDNA was electrophoresed until concentrated near the lane origin. The area of the gel containing the concentrated cDNA was excised and placed in a microfuge tube, and the approximate volume of the gel slice was determined. An aliquot of water (300 µl) approximately three times the volume of the gel slice was added to the tube, and the agarose was melted by heating to 65° C for 15 minutes. Following equilibration of the sample to 42° C, 10 µl of 1 U/µl β-agarase I (New England

Biolabs, Inc.) was added, and the mixture was incubated for 90 minutes to digest the agarose. After incubation, 40 µl of 3 M Na acetate was added to the sample, and the mixture was incubated on ice for 15 minutes. The sample was centrifuged at 14,000 x g for 15 minutes at room temperature to remove undigested agarose. The cDNA in the supernatant was ethanol precipitated, washed in 70% ethanol, air-dried and resuspended in 37 µl of water for the kinase reaction to phosphorylate the ligated *Eco* RI adapters.

To the 37 µl cDNA solution described above was added 10 µl 10x ligase buffer (Stratagene Cloning Systems), and the mixture was heated to 65° C for 5 minutes. The mixture was cooled on ice, and 5 µl 10 mM ATP and 3 µl of 10 U/µl T4 polynucleotide kinase (Stratagene Cloning Systems) were added. The reaction was incubated at 37°C for 45 minutes and was terminated by heating to 65° C for 10 minutes followed by serial extractions with phenol/chloroform and chloroform. The phosphorylated cDNA was ethanol precipitated in the presence of 2.5 M ammonium acetate, washed with 70% ethanol, air dried and resuspended in 12.5 µl water. The concentration of the phosphorylated cDNA was estimated to be ~40 fmole/µl.

The resulting cDNA was cloned into the lambda phage vector λExCell™ (Pharmacia LKB Biotechnology Inc.), purchased predigested with *Eco* RI and *Not* I and dephosphorylated. Ligation of cDNA to vector was carried out in a reaction containing 2 µl of 20 fmole/µl prepared λExCell™ phage arms, 4 µl of water, 1 µl 10x ligase buffer (Promega Corp.), 2 µl of 40 fmole/µl cDNA and 1 µl of 15 U/µl T4 DNA ligase (Promega Corp.). Ligation was carried out at 4° C for 48 hours. Approximately 50% of the ligation mixture was packaged into phage using GIGAPACK® II Gold packaging extract (Stratagene Cloning Systems) according to the directions of the vendor. The resulting cDNA library contained over 1.5×10^7 independent

recombinants with background levels of insertless phage of less than 1.5%.

A ^{32}P-labeled human *MPL-K* receptor cDNA probe was used to isolate mouse *MPL* receptor cDNA from the mouse spleen cDNA phage library. The cDNA library was plated on SURE® strain of *E. coli* cells (Stratagene Cloning Systems) at a density of 40,000 to 50,000 PFU/150 mm diameter plate. Phage plaques from thirty-three plates were transferred onto nylon membranes (Hybond N™; Amersham Corp., Arlington Heights, IL) and processed according to the directions of the manufacturer. The processed filters were baked for 2 hours at 80° C in a vacuum oven followed by several washes at 70° C in wash buffer (0.25 x SSC, 0.25% SDS, 1 mM EDTA) and prehybridized overnight at 65° C in hybridization solution (5x SSC, 5x Denhardt's solution, 0.1% SDS, 1 mM EDTA and 100 µg/ml heat denatured salmon sperm DNA) in a hybridization oven (model HB-2; Techne Inc., Princeton, NJ). Following prehybridization, the hybridization solution was discarded and replaced with fresh hybridization solution containing approximately 2 x 10^6 cpm/ml of ^{32}P-labeled human *MPL-K* cDNA prepared by the use of a commercially available labeling kit (MEGAPRIME™ kit; Amersham Corp., Arlington Heights, IL). The probe was denatured at 98° C for 5 minutes before being added to the hybridization solution. Hybridization was at 65° C overnight. The filters were washed at 55° C in wash buffer (0.25 x SSC, 0.25% SDS, 1 mM EDTA) and were autoradiographed with intensifying screens for 4 days at -70° C on XAR-5 film (Eastman Kodak Co., Rochester, NY). Employing the autoradiograph as template, agar plugs were recovered from regions of the plates corresponding to primary signals and were soaked in SM (0.1 M NaCl; 50 mM Tris-HCl, pH 7.5, 0.02% gelatin) to elute phage for plaque purification. Seven plaque-purified phages were isolated that carried inserts hybridizing to the human *MPL-K* receptor probe. The phagemids contained within the λ

ExCell™ phage were recovered using the *in vivo* recombination system in accordance with the directions of the vendor. The identity of the cDNA inserts was confirmed by DNA sequencing.

The isolated clones encoded a protein exhibiting a high degree of sequence identity to human MPL-P receptor and to a recently reported mouse MPL receptor (Skoda et al., EMBO J. 12: 2645-2653, 1993). The seven clones fell into two classes differing from each other by three clones having a deletion of sequences encoding a stretch of 60 amino acid residues near the N-terminus. The cDNA encoding the protein without the deletion was referred to as mouse Type I MPL receptor cDNA. Type II receptor cDNA lacked sequences encoding Type I receptor residues 131 to 190 of SEQ ID NO:8. In addition, Type I and II receptors differed from the reported mouse MPL receptor sequence (Skoda et al., ibid.) by the presence of a sequence encoding the amino acid residues Val-Arg-Thr-Ser-Pro-Ala-Gly-Glu (SEQ ID NO:9) inserted after amino acid residue 222 and by a substitution of a glycine residue for serine at position 241 (positions refer to the Type I mouse receptor).

Type I and II mouse *MPL* receptor cDNAs were subcloned into the plasmid vector pHZ-1 for expression in mammalian cells. Plasmid pHZ-1 is an expression vector that may be used to express protein in mammalian cells or in a frog oocyte translation system from mRNAs that have been transcribed *in vitro*. The pHZ-1 expression unit comprises the mouse metallothionein-1 promoter, the bacteriophage T7 promoter flanked by multiple cloning banks containing unique restriction sites for insertion of coding sequences, the human growth hormone terminator and the bacteriophage T7 terminator. In addition, pHZ-1 contains an *E. coli* origin of replication; a bacterial beta lactamase gene; a mammalian selectable marker expression unit comprising the SV40 promoter and origin, a

neomycin resistance gene and the SV40 transcription terminator. To facilitate directional cloning into pHZ-1, a polymerase chain reaction employing appropriate primers was used to create an *Eco* RI site upstream from the translation initiation codon and a *Xho* I site downstream from the translation termination codon. The polymerase chain reaction was carried out in a mixture containing 10 µl 10x ULTMA™ DNA polymerase buffer (Roche Molecular Systems, Inc., Branchburg, NJ), 6 µl of 25 mM $MgCl_2$, 0.2 µl of a deoxynucleotide triphosphate solution containing 10 mM each of dATP, dGTP, dTTP and dCTP (Pharmacia LKB Biotechnology Inc.), 2.5 µl of 20 pmole/µl primer ZC6603 (SEQ ID NO:10), 2.5 µl of 20 pmole/µl primer ZC5762 (SEQ ID NO:11), 32.8 µl of water, 1 µl of an early log phase bacterial culture harboring either a Type I or a Type II mouse MPL receptor plasmid and 1 µl of 6 U/µl DNA polymerase (ULTMA™ polymerase; Roche Molecular Systems, Inc., Branchburg, NJ). AmpliWax™ (Roche Molecular Systems, Inc.) was employed in the reaction according to the directions of the vendor. The polymerase chain reaction was run for 25 cycles (1 minute at 95° C, 1 minute at 55° C and 3 minutes at 72° C) followed by a 10 minute incubation at 72° C. The amplified products were serially extracted with phenol/chloroform and chloroform, then ethanol precipitated in the presence of 6 µg glycogen carrier and 2.5 M ammonium acetate. The pellets were resuspended in 87 µl of water to which was added 10 µl of 10 x H buffer (Boehringer Mannheim, Inc.), 2 µl of 10 U/µl *Eco* RI (Boehringer Mannheim, Inc.) and 1 µl of 40 U/µl *Xho* I (Boehringer Mannheim, Inc.). Digestion was carried out at 37° C for 1 hour. The reaction was terminated by heating to 65° C for 15 minutes and chromatographed through a 400 pore size gel filtration column (CHROMA SPIN + TE-400™; Clontech Laboratories Inc.).

The isolated receptor inserts described above were ligated into *Eco* RI and *Xho* I digested and

dephosphorylated pHZ-1 vector. The ligation reaction contained 1 µl of 50 ng/µl prepared pHZ-1 vector, 5 µl of 5 ng/µl cDNA insert, 2 µl of 10x ligase buffer (Promega Corp.), 11.75 µl water and 0.25 µl of 4 U/µl T4 DNA ligase (Stratagene Cloning Systems). Ligation was carried out at 10° C overnight. The ligated DNAs were transfected into *E. coli* (MAX EFFICIENCY DH10B™ competent cells; GIBCO BRL) in accordance with the vendor's directions. The validity of Type I and Type II mouse *MPL* and human *MPL-P* receptor inserts in pHZ-1 was confirmed by DNA sequencing. The resulting plasmids pSLmpl-8 and pSLmpl-9 carried the mouse Type II and Type I *MPL* receptor cDNAs, respectively. Plasmid pSLmpl-44 carried the human *MPL-P* cDNA insert.

A mammalian expression plasmid encoding soluble mouse Type I MPL receptor (pLDmpl-53) was produced by combining DNA segments from pSLmpl-9, a mammalian expression plasmid containing the cDNA encoding full-length mouse Type I *MPL* receptor described above, with a DNA segment from pSLmpl-26, an expression plasmid constructed to produce the soluble mouse Type I MPL receptor in bacteria.

A cDNA segment encoding mouse Type I MPL soluble receptor was isolated by PCR employing primers ZC6704 (SEQ ID NO:12) and ZC6703 (SEQ ID NO:13) using full-length receptor plasmid pSLmpl-9 as template. To facilitate directional cloning, primers ZC6704 and ZC6703 incorporated *Eco* RI and *Xho* I restriction sites at their respective 5' ends. Primer ZC6703 also encoded an inframe consensus target sequence for protein kinase to enable *in vitro* labeling of the purified soluble receptor with ^{32}P γ-ATP (Li et al., Proc. Natl. Acad. Sci. U.S.A. 86: 558-562, 1989). The PCR was carried out in a mixture containing 10 µl 10x ULTMA™ DNA polymerase buffer (Roche Molecular Systems, Inc.), 6 µl of 25 mM $MgCl_2$, 0.2 µl of a deoxynucleotide triphosphate solution containing 10 mM each of dATP, dGTP, dTTP and dCTP (Pharmacia LKB

Biotechnology Inc.), 11 µl of 4.55 pmole/µl primer ZC6704 (SEQ ID NO:12), 21 µl of 2.43 pmole/µl primer ZC6703 (SEQ ID NO:13), 50.3 µl of water, 1 µl 50 ng/µl *Hind* III and *Xba* I digested pSLmpl-9 and 1 µl of 6 U/µl ULTMA™ DNA polymerase (Roche Molecular Systems, Inc.). AmpliWax™ (Roche Molecular Systems, Inc.) was employed in the reaction according to the directions of the vendor. The polymerase chain reaction was run for 3 cycles (1 minute at 95° C, 1 minute at 50° C and 2 minutes at 72° C) followed by 11 cycles at increased hybridization stringency (1 minute at 95° C, 30 seconds at 55° C and 2 minutes at 72° C) followed by a 10 minute incubation at 72° C. The amplified product was serially extracted with phenol/chloroform and chloroform followed by chromatography through a 400 pore size gel filtration column (Clontech Laboratories, Inc.). The PCR product was ethanol precipitated in the presence of 20 µg glycogen carrier and 2.5 M ammonium acetate. The pellet was resuspended in 32 µl of water. To 16 µl of the resuspended PCR product was added 2 µl 10x H buffer (Boehringer Mannheim, Inc.), 1 µl of 10 U/µl *Eco* RI (Boehringer Mannheim, Inc.) and 1 µl of 40 U/µl *Xho* I (Boehringer Mannheim, Inc.). Digestion was carried out at 37° C for 1 hour. Digestion was terminated by heating to 65° C for 15 minutes, and DNA was purified on a 0.7% low-melt agarose gel. Fragment recovery from low-melt agarose was done by digestion of the gel matrix with β-agarase I (New England Biolabs).

The resulting PCR product encoded the N-terminal extracellular domain of mouse Type I MPL receptor (residues 27 to 480 of SEQ ID NO:8). In the absence of the putative receptor trans-membrane domain (residues 483 to 504 of SEQ ID NO:8) the expressed protein is expected to be secreted in the presence of a suitable signal peptide. A mouse Type II soluble *MPL* receptor encoding cDNA was obtained using the PCR conditions described above

except that pSLmpl-8 was used as template. The validity of both receptor fragments was confirmed by DNA sequencing.

The soluble mouse Type I and Type II *MPL* receptor encoding DNA fragments were cloned into *Eco RI* and *Xho* I digested vector pOmpA2-5 to yield pSLmpl-26 and pSLmpl-27, respectively. Plasmid pOmpA2-5 is a modification of pOmpA2 (Ghrayab et al., EMBO J. 3: 2437-2442, 1984), a bacterial expression vector designed to target the recombinant protein to the periplasmic space. pOmpA2-5 was constructed by replacement of a 13 bp sequence between the *Eco* RI and *Bam* HI sites of pOmpA2 with a synthetic 42 bp sequence. The sequence was created by annealing of two 42 nucleotide complementary oligonucleotides (ZC6707, SEQ ID NO:14; ZC 6706, SEQ ID NO:15), which when base paired formed *Eco* RI and *Bam* HI cohesive ends, facilitating directional cloning into *Eco* RI and *Bam* HI digested pOmpA2. Within the inserted sequence is an *Xho* I site inframed with respect to a bacterial leader sequence and to the mouse *MPL* soluble receptor encoding cDNAs described above, as well as an inframe tract of 6 histidine codons located 3' of the *Xho* I site to enable the recombinant protein to be purified by metal chelation affinity chromatography (Houchuli et al., Bio/Technol. 6: 1321-1325, 1988). Following the sequence encoding the histidine tract was an inframe termination codon. The validity of the pOmpA2-5, pSLmpl-26 and pSLmpl-27 was confirmed by DNA sequencing.

pLDmpl-53, a mammalian expression plasmid producing soluble mouse Type I MPL receptor, was constructed by combining DNA segments from pSLmpl-9 and pSLmpl-26 into expression vector pHZ-200 (pHZ-1 in which a dihydrofolate reductase sequence was substituted for the neomycin resistance gene). The 1164 bp *Eco* RI/*Bam* HI cDNA fragment from pSLmpl-9 replaced the mammalian signal sequence deleted during the construction of bacterial

expression plasmid pSLmpl-26. The 416 bp *Bam* HI fragment from pSLmpl-26 supplied the coding sequence for the carboxy-terminal portion of the soluble MPL receptor, the kinase labeling domain, the poly-histidine tract and the translation terminator. The two fragments were gel purified and cloned into the *Eco* RI/*Bam* HI sites of pBluescript® KS+ (Stratagene Cloning Systems) to yield plasmid pBS8.76LD-5. Correct orientation of the 416 bp pSLmpl-26 derived *Bam* HI fragment with respect to the 1164 bp pSLmpl-9 derived *Eco* RI/*Bam* HI fragment in pBS8.76LD-5 was determined by PCR using primers ZC 6603 (SEQ ID NO:10) and ZC 6703 (SEQ ID NO:13). The *Xba* I site within the poly-linker sequence of pBS8.76LD-5 enabled the reconstituted receptor cDNA to be excised as a 1.5 kb *Eco* RI/*Xba* I fragment for cloning into pHZ-200 following digestion of the vector with *Eco* RI and *Xba* I. The resulting mammalian expression plasmid, pLDmpl-53, was prepared in large scale for transfection into BHK cells.

Twenty micrograms of purified pLDmpl-53 plasmid was transfected into BHK 570 cells using the calcium phosphate precipitation method. After 5 hours, the cells were shocked with 15% glycerol for 3 minutes to facilitate uptake of DNA. Fresh growth media was added overnight. The following day the cells were split at various dilutions, and selection media containing 1 μM methotrexate was added. After approximately two weeks, discrete, methotrexate-resistant colonies were visible. Resistant colonies were either pooled or maintained as distinct clones. Spent media from the pooled colonies was immediately tested for presence of soluble MPL receptor protein.

Soluble MPL receptor protein was isolated through the interaction of the poly-histidine tract present on the carboxy-terminal of the protein with a metal chelation resin containing immobilized Ni^{2+} (HIS-BIND™; Novagen, Madison, WI). Serum-free spent culture

media from the pLDmpl-53 pool was passed over the resin, and bound protein was eluted with 1 M imidazole. SDS-PAGE analysis revealed a single band at ~67 kDa. This protein was subjected to N-terminal amino acid analysis and confirmed to be mouse MPL receptor.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: ZymoGenetics, Inc.
1201 Eastlake Avenue East
Seattle
WA
USA
98102

(ii) TITLE OF INVENTION: LOW MOLECULAR WEIGHT THROMBOPOIETIN

(iii) NUMBER OF SEQUENCES: 15

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: ZymoGenetics, Inc.
(B) STREET: 1201 Eastlake Avenue East
(C) CITY: Seattle
(D) STATE: WA
(E) COUNTRY: USA
(F) ZIP: 98102

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Parker, Gary E
(B) REGISTRATION NUMBER: 31-648
(C) REFERENCE/DOCKET NUMBER: 94-12PC

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 206-442-6673
(B) TELEFAX: 206-442-6678

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1486 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 105..1241

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CCTCGTGCCG GTCCTGAGGC CCTTCTCCAC CCGGACAGAG TCCTTGGCCC ACCTCTCTCC      60

CACCCGACTC TGCCGAAAGA AGCACAGAAG CTCAAGCCGC CTCC ATG GCC CCA GGA      116
                                                 Met Ala Pro Gly
                                                   1

AAG ATT CAG GGG AGA GGC CCC ATA CAG GGA GCC ACT TCA GTT AGA CAC      164
Lys Ile Gln Gly Arg Gly Pro Ile Gln Gly Ala Thr Ser Val Arg His
  5                  10                  15                  20

CTG GCC AGA ATG GAG CTG ACT GAT TTG CTC CTG GCG GCC ATG CTT CTT      212
Leu Ala Arg Met Glu Leu Thr Asp Leu Leu Leu Ala Ala Met Leu Leu
              25                  30                  35

GCA GTG GCA AGA CTA ACT CTG TCC AGC CCC GTA GCT CCT GCC TGT GAC      260
Ala Val Ala Arg Leu Thr Leu Ser Ser Pro Val Ala Pro Ala Cys Asp
          40                  45                  50

CCC AGA CTC CTA AAT AAA CTG CTG CGT GAC TCC CAC CTC CTT CAC AGC      308
Pro Arg Leu Leu Asn Lys Leu Leu Arg Asp Ser His Leu Leu His Ser
      55                  60                  65

CGA CTG AGT CAG TGT CCC GAC GTC GAC CCT TTG TCT ATC CCT GTT CTG      356
Arg Leu Ser Gln Cys Pro Asp Val Asp Pro Leu Ser Ile Pro Val Leu
  70                  75                  80
```

```
CTG CCT GCT GTG GAC TTT AGC CTG GGA GAA TGG AAA ACC CAG ACG GAA      404
Leu Pro Ala Val Asp Phe Ser Leu Gly Glu Trp Lys Thr Gln Thr Glu
 85                  90                  95                 100

CAG AGC AAG GCA CAG GAC ATT CTA GGG GCA GTG TCC CTT CTA CTG GAG      452
Gln Ser Lys Ala Gln Asp Ile Leu Gly Ala Val Ser Leu Leu Leu Glu
                105                 110                 115

GGA GTG ATG GCA GCA CGA GGA CAG TTG GAA CCC TCC TGC CTC TCA TCC      500
Gly Val Met Ala Ala Arg Gly Gln Leu Glu Pro Ser Cys Leu Ser Ser
            120                 125                 130

CTC CTG GGA CAG CTT TCT GGG CAG GTT CGC CTC CTC TTG GGG GCC CTG      548
Leu Leu Gly Gln Leu Ser Gly Gln Val Arg Leu Leu Leu Gly Ala Leu
        135                 140                 145

CAG GGC CTC CTA GGA ACC CAG CTT CCT CTA CAG GGC AGG ACC ACA GCT      596
Gln Gly Leu Leu Gly Thr Gln Leu Pro Leu Gln Gly Arg Thr Thr Ala
    150                 155                 160

CAC AAG GAC CCC AAT GCC CTC TTC TTG AGC TTG CAA CAA CTG CTT CGG      644
His Lys Asp Pro Asn Ala Leu Phe Leu Ser Leu Gln Gln Leu Leu Arg
165                 170                 175                 180

GGA AAG GTG CGC TTC CTG CTT CTG GTA GAA GGT CCC ACC CTC TGT GTC      692
Gly Lys Val Arg Phe Leu Leu Leu Val Glu Gly Pro Thr Leu Cys Val
                185                 190                 195

AGA CGG ACC CTG CCA ACC ACA GCT GTC CCA AGC AGT ACT TCT CAA CTC      740
Arg Arg Thr Leu Pro Thr Thr Ala Val Pro Ser Ser Thr Ser Gln Leu
            200                 205                 210

CTC ACA CTA AAC AAG TTC CCA AAC AGG ACT TCT GGA TTG TTG GAG ACG      788
Leu Thr Leu Asn Lys Phe Pro Asn Arg Thr Ser Gly Leu Leu Glu Thr
        215                 220                 225

AAC TTC AGT GTC ACA GCC AGA ACT GCT GGC CCT GGA CTT CTG AGC AGG      836
Asn Phe Ser Val Thr Ala Arg Thr Ala Gly Pro Gly Leu Leu Ser Arg
    230                 235                 240

CTT CAG GGA TTC AGA GTC AAG ATT ACT CCT GGT CAG CTA AAT CAA ACC      884
Leu Gln Gly Phe Arg Val Lys Ile Thr Pro Gly Gln Leu Asn Gln Thr
245                 250                 255                 260
```

```
TCC AGG TCC CCA GTC CAA ATC TCT GGA TAC CTG AAC AGG ACA CAC GGA      932
Ser Arg Ser Pro Val Gln Ile Ser Gly Tyr Leu Asn Arg Thr His Gly
                265                 270                 275

CCT GTG AAT GGA ACT CAT GGG CTC TTT GCT GGA ACC TCA CTT CAG ACC      980
Pro Val Asn Gly Thr His Gly Leu Phe Ala Gly Thr Ser Leu Gln Thr
            280                 285                 290

CTG GAA GCC TCA GAC ATC TCG CCC GGA GCT TTC AAC AAA GGC TCC CTG     1028
Leu Glu Ala Ser Asp Ile Ser Pro Gly Ala Phe Asn Lys Gly Ser Leu
        295                 300                 305

GCA TTC AAC CTC CAG GGT GGA CTT CCT CCT TCT CCA AGC CTT GCT CCT     1076
Ala Phe Asn Leu Gln Gly Gly Leu Pro Pro Ser Pro Ser Leu Ala Pro
    310                 315                 320

GAT GGA CAC ACA CCC TTC CCT CCT TCA CCT GCC TTG CCC ACC ACC CAT     1124
Asp Gly His Thr Pro Phe Pro Pro Ser Pro Ala Leu Pro Thr Thr His
325                 330                 335                 340

GGA TCT CCA CCC CAG CTC CAC CCC CTG TTT CCT GAC CCT TCC ACC ACC     1172
Gly Ser Pro Pro Gln Leu His Pro Leu Phe Pro Asp Pro Ser Thr Thr
                345                 350                 355

ATG CCT AAC TCT ACC GCC CCT CAT CCA GTC ACA ATG TAC CCT CAT CCC     1220
Met Pro Asn Ser Thr Ala Pro His Pro Val Thr Met Tyr Pro His Pro
            360                 365                 370

AGG AAT TTG TCT CAG GAA ACA TAGCGCGGGC ACTGGCCCAG TGAGCGTCTG       1271
Arg Asn Leu Ser Gln Glu Thr
        375

CAGCTTCTCT CGGGGACAAG CTTCCCCAGG AAGGCTGAGA GGCAGCTGCA TCTGCTCCAG  1331

ATGTTCTGCT TTCACCTAAA AGGCCCTGGG GAAGGGATAC ACAGCACTGG AGATTGTAAA  1391

ATTTTAGGAG CTATTTTTTT TTAACCTATC AGCAATATTC ATCAGAGCAG CTAGCGATCT  1451

TTGGTCTATT TTCGGTATAA ATTTGAAAAT CACTA                              1486
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 379 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala Pro Gly Lys Ile Gln Gly Arg Gly Pro Ile Gln Gly Ala Thr
1 5 10 15

Ser Val Arg His Leu Ala Arg Met Glu Leu Thr Asp Leu Leu Leu Ala
20 25 30

Ala Met Leu Leu Ala Val Ala Arg Leu Thr Leu Ser Ser Pro Val Ala
35 40 45

Pro Ala Cys Asp Pro Arg Leu Leu Asn Lys Leu Leu Arg Asp Ser His
50 55 60

Leu Leu His Ser Arg Leu Ser Gln Cys Pro Asp Val Asp Pro Leu Ser
65 70 75 80

Ile Pro Val Leu Leu Pro Ala Val Asp Phe Ser Leu Gly Glu Trp Lys
85 90 95

Thr Gln Thr Glu Gln Ser Lys Ala Gln Asp Ile Leu Gly Ala Val Ser
100 105 110

Leu Leu Leu Glu Gly Val Met Ala Ala Arg Gly Gln Leu Glu Pro Ser
115 120 125

Cys Leu Ser Ser Leu Leu Gly Gln Leu Ser Gly Gln Val Arg Leu Leu
130 135 140

Leu Gly Ala Leu Gln Gly Leu Leu Gly Thr Gln Leu Pro Leu Gln Gly
145 150 155 160

Arg Thr Thr Ala His Lys Asp Pro Asn Ala Leu Phe Leu Ser Leu Gln
165 170 175

Gln Leu Leu Arg Gly Lys Val Arg Phe Leu Leu Leu Val Glu Gly Pro
180 185 190

Thr Leu Cys Val Arg Arg Thr Leu Pro Thr Thr Ala Val Pro Ser Ser
195 200 205

```
Thr Ser Gln Leu Leu Thr Leu Asn Lys Phe Pro Asn Arg Thr Ser Gly
    210                 215                 220

Leu Leu Glu Thr Asn Phe Ser Val Thr Ala Arg Thr Ala Gly Pro Gly
225                 230                 235                 240

Leu Leu Ser Arg Leu Gln Gly Phe Arg Val Lys Ile Thr Pro Gly Gln
                245                 250                 255

Leu Asn Gln Thr Ser Arg Ser Pro Val Gln Ile Ser Gly Tyr Leu Asn
            260                 265                 270

Arg Thr His Gly Pro Val Asn Gly Thr His Gly Leu Phe Ala Gly Thr
        275                 280                 285

Ser Leu Gln Thr Leu Glu Ala Ser Asp Ile Ser Pro Gly Ala Phe Asn
    290                 295                 300

Lys Gly Ser Leu Ala Phe Asn Leu Gln Gly Gly Leu Pro Pro Ser Pro
305                 310                 315                 320

Ser Leu Ala Pro Asp Gly His Thr Pro Phe Pro Pro Ser Pro Ala Leu
                325                 330                 335

Pro Thr Thr His Gly Ser Pro Pro Gln Leu His Pro Leu Phe Pro Asp
            340                 345                 350

Pro Ser Thr Thr Met Pro Asn Ser Thr Ala Pro His Pro Val Thr Met
        355                 360                 365

Tyr Pro His Pro Arg Asn Leu Ser Gln Glu Thr
    370                 375
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1062 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1059

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATG GAG CTG ACT GAA TTG CTC CTC GTG GTC ATG CTT CTC CTA ACT GCA          48
Met Glu Leu Thr Glu Leu Leu Leu Val Val Met Leu Leu Leu Thr Ala
 1               5                  10                  15

AGG CTA ACG CTG TCC AGC CCG GCT CCT CCT GCT TGT GAC CTC CGA GTC          96
Arg Leu Thr Leu Ser Ser Pro Ala Pro Pro Ala Cys Asp Leu Arg Val
            20                  25                  30

CTC AGT AAA CTG CTT CGT GAC TCC CAT GTC CTT CAC AGC AGA CTG AGC         144
Leu Ser Lys Leu Leu Arg Asp Ser His Val Leu His Ser Arg Leu Ser
        35                  40                  45

CAG TGC CCA GAG GTT CAC CCT TTG CCT ACA CCT GTC CTG CTG CCT GCT         192
Gln Cys Pro Glu Val His Pro Leu Pro Thr Pro Val Leu Leu Pro Ala
    50                  55                  60

GTG GAC TTT AGC TTG GGA GAA TGG AAA ACC CAG ATG GAG GAG ACC AAG         240
Val Asp Phe Ser Leu Gly Glu Trp Lys Thr Gln Met Glu Glu Thr Lys
65                  70                  75                  80

GCA CAG GAC ATT CTG GGA GCA GTG ACC CTT CTG CTG GAG GGA GTG ATG         288
Ala Gln Asp Ile Leu Gly Ala Val Thr Leu Leu Leu Glu Gly Val Met
                85                  90                  95

GCA GCA CGG GGA CAA CTG GGA CCC ACT TGC CTC TCA TCC CTC CTG GGG         336
Ala Ala Arg Gly Gln Leu Gly Pro Thr Cys Leu Ser Ser Leu Leu Gly
            100                 105                 110

CAG CTT TCT GGA CAG GTC CGT CTC CTC CTT GGG GCC CTG CAG AGC CTC         384
Gln Leu Ser Gly Gln Val Arg Leu Leu Leu Gly Ala Leu Gln Ser Leu
        115                 120                 125

CTT GGA ACC CAG CTT CCT CCA CAG GGC AGG ACC ACA GCT CAC AAG GAT         432
Leu Gly Thr Gln Leu Pro Pro Gln Gly Arg Thr Thr Ala His Lys Asp
    130                 135                 140
```

```
CCC AAT GCC ATC TTC CTG AGC TTC CAA CAC CTG CTC CGA GGA AAG GTG        480
Pro Asn Ala Ile Phe Leu Ser Phe Gln His Leu Leu Arg Gly Lys Val
145                 150                 155                 160

CGT TTC CTG ATG CTT GTA GGA GGG TCC ACC CTC TGC GTC AGG CGG GCC        528
Arg Phe Leu Met Leu Val Gly Gly Ser Thr Leu Cys Val Arg Arg Ala
                165                 170                 175

CCA CCC ACC ACA GCT GTC CCC AGC AGA ACC TCT CTA GTC CTC ACA CTG        576
Pro Pro Thr Thr Ala Val Pro Ser Arg Thr Ser Leu Val Leu Thr Leu
            180                 185                 190

AAC GAG CTC CCA AAC AGG ACT TCT GGA TTG TTG GAG ACA AAC TTC ACT        624
Asn Glu Leu Pro Asn Arg Thr Ser Gly Leu Leu Glu Thr Asn Phe Thr
        195                 200                 205

GCC TCA GCC AGA ACT ACT GGC TCT GGG CTT CTG AAG TGG CAG CAG GGA        672
Ala Ser Ala Arg Thr Thr Gly Ser Gly Leu Leu Lys Trp Gln Gln Gly
    210                 215                 220

TTC AGA GCC AAG ATT CCT GGT CTG CTG AAC CAA ACC TCC AGG TCC CTG        720
Phe Arg Ala Lys Ile Pro Gly Leu Leu Asn Gln Thr Ser Arg Ser Leu
225                 230                 235                 240

GAC CAA ATC CCC GGA TAC CTG AAC AGG ATA CAC GAA CTC TTG AAT GGA        768
Asp Gln Ile Pro Gly Tyr Leu Asn Arg Ile His Glu Leu Leu Asn Gly
                245                 250                 255

ACT CGT GGA CTC TTT CCT GGA CCC TCA CGC AGG ACC CTA GGA GCC CCG        816
Thr Arg Gly Leu Phe Pro Gly Pro Ser Arg Arg Thr Leu Gly Ala Pro
            260                 265                 270

GAC ATT TCC TCA GGA ACA TCA GAC ACA GGC TCC CTG CCA CCC AAC CTC        864
Asp Ile Ser Ser Gly Thr Ser Asp Thr Gly Ser Leu Pro Pro Asn Leu
        275                 280                 285

CAG CCT GGA TAT TCT CCT TCC CCA ACC CAT CCT CCT ACT GGA CAG TAT        912
Gln Pro Gly Tyr Ser Pro Ser Pro Thr His Pro Pro Thr Gly Gln Tyr
    290                 295                 300

ACG CTC TTC CCT CTT CCA CCC ACC TTG CCC ACC CCT GTG GTC CAG CTC        960
Thr Leu Phe Pro Leu Pro Pro Thr Leu Pro Thr Pro Val Val Gln Leu
305                 310                 315                 320
```

```
CAC CCC CTG CTT CCT GAC CCT TCT GCT CCA ACG CCC ACC CCT ACC AGC      1008
His Pro Leu Leu Pro Asp Pro Ser Ala Pro Thr Pro Thr Pro Thr Ser
                325                 330                 335

CCT CTT CTA AAC ACA TCC TAC ACC CAC TCC CAG AAT CTG TCT CAG GAA      1056
Pro Leu Leu Asn Thr Ser Tyr Thr His Ser Gln Asn Leu Ser Gln Glu
            340                 345                 350

GGG TAA                                                              1062
Gly
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 353 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Glu Leu Thr Glu Leu Leu Leu Val Val Met Leu Leu Leu Thr Ala
 1               5                  10                  15

Arg Leu Thr Leu Ser Ser Pro Ala Pro Pro Ala Cys Asp Leu Arg Val
            20                  25                  30

Leu Ser Lys Leu Leu Arg Asp Ser His Val Leu His Ser Arg Leu Ser
        35                  40                  45

Gln Cys Pro Glu Val His Pro Leu Pro Thr Pro Val Leu Leu Pro Ala
    50                  55                  60

Val Asp Phe Ser Leu Gly Glu Trp Lys Thr Gln Met Glu Glu Thr Lys
65                  70                  75                  80

Ala Gln Asp Ile Leu Gly Ala Val Thr Leu Leu Leu Glu Gly Val Met
                85                  90                  95

Ala Ala Arg Gly Gln Leu Gly Pro Thr Cys Leu Ser Ser Leu Leu Gly
            100                 105                 110
```

Gln Leu Ser Gly Gln Val Arg Leu Leu Leu Gly Ala Leu Gln Ser Leu
115 120 125

Leu Gly Thr Gln Leu Pro Pro Gln Gly Arg Thr Thr Ala His Lys Asp
130 135 140

Pro Asn Ala Ile Phe Leu Ser Phe Gln His Leu Leu Arg Gly Lys Val
145 150 155 160

Arg Phe Leu Met Leu Val Gly Gly Ser Thr Leu Cys Val Arg Arg Ala
165 170 175

Pro Pro Thr Thr Ala Val Pro Ser Arg Thr Ser Leu Val Leu Thr Leu
180 185 190

Asn Glu Leu Pro Asn Arg Thr Ser Gly Leu Leu Glu Thr Asn Phe Thr
195 200 205

Ala Ser Ala Arg Thr Thr Gly Ser Gly Leu Leu Lys Trp Gln Gln Gly
210 215 220

Phe Arg Ala Lys Ile Pro Gly Leu Leu Asn Gln Thr Ser Arg Ser Leu
225 230 235 240

Asp Gln Ile Pro Gly Tyr Leu Asn Arg Ile His Glu Leu Leu Asn Gly
245 250 255

Thr Arg Gly Leu Phe Pro Gly Pro Ser Arg Arg Thr Leu Gly Ala Pro
260 265 270

Asp Ile Ser Ser Gly Thr Ser Asp Thr Gly Ser Leu Pro Pro Asn Leu
275 280 285

Gln Pro Gly Tyr Ser Pro Ser Pro Thr His Pro Pro Thr Gly Gln Tyr
290 295 300

Thr Leu Phe Pro Leu Pro Pro Thr Leu Pro Thr Pro Val Val Gln Leu
305 310 315 320

His Pro Leu Leu Pro Asp Pro Ser Ala Pro Thr Pro Thr Pro Thr Ser
325 330 335

Pro Leu Leu Asn Thr Ser Tyr Thr His Ser Gln Asn Leu Ser Gln Glu
340 345 350

Gly

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(B) CLONE: ZC8045

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AGAAGTCCTG TTTGAATTCT AGTTCAGTGT GAGGACATTA AGA 43

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(B) CLONE: ZC7878

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCGTCTCCTC CTTGGGGCCC ATTAAGA 27

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(B) CLONE: ZC6091

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GAGCACAGAA TTCACTACTC GAGGCGGCCG CTTTTTTTTT TTTTTTTTT 49

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 633 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Pro Ser Trp Ala Leu Phe Met Val Thr Ser Cys Leu Leu Leu Ala
1 5 10 15

Leu Pro Asn Gln Ala Gln Val Thr Ser Gln Asp Val Phe Leu Leu Ala
20 25 30

Leu Gly Thr Glu Pro Leu Asn Cys Phe Ser Gln Thr Phe Glu Asp Leu
35 40 45

Thr Cys Phe Trp Asp Glu Glu Glu Ala Ala Pro Ser Gly Thr Tyr Gln
50 55 60

Leu Leu Tyr Ala Tyr Arg Gly Glu Lys Pro Arg Ala Cys Pro Leu Tyr
65 70 75 80

Ser Gln Ser Val Pro Thr Phe Gly Thr Arg Tyr Val Cys Gln Phe Pro
85 90 95

Ala Gln Asp Glu Val Arg Leu Phe Phe Pro Leu His Leu Trp Val Lys
100 105 110

Asn Val Ser Leu Asn Gln Thr Leu Ile Gln Arg Val Leu Phe Val Asp
115 120 125

Ser Val Gly Leu Pro Ala Pro Pro Arg Val Ile Lys Ala Arg Gly Gly
130 135 140

Ser Gln Pro Gly Glu Leu Gln Ile His Trp Glu Ala Pro Ala Pro Glu
145 150 155 160

Ile Ser Asp Phe Leu Arg His Glu Leu Arg Tyr Gly Pro Thr Asp Ser
165 170 175

Ser Asn Ala Thr Ala Pro Ser Val Ile Gln Leu Leu Ser Thr Glu Thr
180 185 190

Cys Cys Pro Thr Leu Trp Met Pro Asn Pro Val Pro Val Leu Asp Gln
195 200 205

Pro Pro Cys Val His Pro Thr Ala Ser Gln Pro His Gly Pro Val Arg
210 215 220

Thr Ser Pro Ala Gly Glu Ala Pro Phe Leu Thr Val Lys Gly Gly Ser
225 230 235 240

Cys Leu Val Ser Gly Leu Gln Ala Gly Lys Ser Tyr Trp Leu Gln Leu
245 250 255

Arg Ser Gln Pro Asp Gly Val Ser Leu Arg Gly Ser Trp Gly Pro Trp
260 265 270

Ser Phe Pro Val Thr Val Asp Leu Pro Gly Asp Ala Val Thr Ile Gly
275 280 285

Leu Gln Cys Phe Thr Leu Asp Leu Lys Met Val Thr Cys Gln Trp Gln
290 295 300

Gln Gln Asp Arg Thr Ser Ser Gln Gly Phe Phe Arg His Ser Arg Thr
305 310 315 320

Arg Cys Cys Pro Thr Asp Arg Asp Pro Thr Trp Glu Lys Cys Glu Glu
325 330 335

Glu Glu Pro Arg Pro Gly Ser Gln Pro Ala Leu Val Ser Arg Cys His
340 345 350

Phe Lys Ser Arg Asn Asp Ser Val Ile His Ile Leu Val Glu Val Thr
355 360 365

Thr Ala Gln Gly Ala Val His Ser Tyr Leu Gly Ser Pro Phe Trp Ile
370 375 380

His Gln Ala Val Leu Leu Pro Thr Pro Ser Leu His Trp Arg Glu Val
385 390 395 400

Ser Ser Gly Arg Leu Glu Leu Glu Trp Gln His Gln Ser Ser Trp Ala
405 410 415

Ala Gln Glu Thr Cys Tyr Gln Leu Arg Tyr Thr Gly Glu Gly Arg Glu
420 425 430

Asp Trp Lys Val Leu Glu Pro Ser Leu Gly Ala Arg Gly Gly Thr Leu
435 440 445

Glu Leu Arg Pro Arg Ala Arg Tyr Ser Leu Gln Leu Arg Ala Arg Leu
450 455 460

Asn Gly Pro Thr Tyr Gln Gly Pro Trp Ser Ala Trp Ser Pro Pro Ala
465 470 475 480

Arg Val Ser Thr Gly Ser Glu Thr Ala Trp Ile Thr Leu Val Thr Ala
485 490 495

Leu Leu Leu Val Leu Ser Leu Ser Ala Leu Leu Gly Leu Leu Leu Leu
500 505 510

Lys Trp Gln Phe Pro Ala His Tyr Arg Arg Leu Arg His Ala Leu Trp
515 520 525

Pro Ser Leu Pro Asp Leu His Arg Val Leu Gly Gln Tyr Leu Arg Asp
530 535 540

Thr Ala Ala Leu Ser Pro Ser Lys Ala Thr Val Thr Asp Ser Cys Glu
545 550 555 560

Glu Val Glu Pro Ser Leu Leu Glu Ile Leu Pro Lys Ser Ser Glu Ser
565 570 575

Thr Pro Leu Pro Leu Cys Pro Ser Gln Pro Gln Met Asp Tyr Arg Gly
580 585 590

Leu Gln Pro Cys Leu Arg Thr Met Pro Leu Ser Val Cys Pro Pro Met
595 600 605

Ala Glu Thr Gly Ser Cys Cys Thr Thr His Ile Ala Asn His Ser Tyr
610 615 620

Leu Pro Leu Ser Tyr Trp Gln Gln Pro
625 630

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Val Arg Thr Ser Pro Ala Gly Glu
1 5

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(B) CLONE: ZC6603

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GAGGAATTCG CAGAAGCCAT GCCCTCTTGG GCCCTCTTCA TGGTC 45

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(B) CLONE: ZC5762

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AGAGAGAGAG AGAGCTCGAG TCAAGGCTGC TGCCAATAGC TTAGTGGTAG GT 52

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(B) CLONE: ZC6704

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GAAGAGGAAT TCACCATGGA TGTCTTCTTG CTGGCCTTGG GCACAGAG 48

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(B) CLONE: ZC6703

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CGACTTTACC TCGAGTGCTA CTGATGCTCT TCTGCCAGCA GTCTCGGAGC CCGTGGACAC 60

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(B) CLONE: ZC6707

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AATTCGCCAT GGGACTCGAG CATCACCATC ACCATCACTG AG 42

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(B) CLONE: ZC6706

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GATCCTCAGT GATGGTGATG GTGATGCTCG AGTCCCATGG CG 42

Claims

We claim:

1. An isolated mammalian thrombopoietin characterized by an amino acid sequence selected from the group consisting of:

(a) a sequence of amino acids as shown in SEQ ID NO:2 having an amino terminus at Ser(45) and a carboxyl terminus between Ser(208) and Asn(216), inclusive;

(b) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus between Arg(185) and Asn(193), inclusive;

(c) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Arg(198);

(d) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Phe(207); and

(e) allelic variants of (a)-(d).

2. An isolated mammalian thrombopoietin according to claim 1, further characterized by:

a) a peptide backbone mass of 17,490 to 19,045 amu as determined by mass spectrometry;

b) absence of N-linked carbohydrate;

c) presence of O-linked carbohydrate; and

d) essentially free of thrombopoietin species having a peptide backbone mass greater than 19045 amu as determined by mass spectrometry.

3. An isolated mammalian thrombopoietin according to claim 1 having an amino acid sequence consisting of:

(a) the sequence shown in SEQ ID NO:2 from an amino-terminus at Ser, residue number 45, to a carboxyl-

terminus between Ser, residue number 208, and Asn, residue number 216, inclusive; or

(b) a sequence which is an allelic variant of (a).

4. An isolated mammalian thrombopoietin according to claim 1 characterized by:

(a) an amino acid sequence as shown in SEQ ID NO:4 from an amino-terminus at Ser, residue number 22, to a carboxyl-terminus between Arg, residue number 185, and Asn, residue number 193, inclusive; or

(b) an amino acid sequence which is an allelic variant of (a).

5. A purified mammalian thrombopoietin according to claim 1 which is a mixture of polypeptides having different carboxyl termini.

6. A composition of mammalian thrombopoietin consisting essentially of one or more polypeptides having an amino acid sequence consisting of:

(a) the sequence shown in SEQ ID NO:4 from an amino-terminus at Ser, residue number 22, to a carboxyl-terminus between Arg, residue number 185, and Asn, residue number 193, inclusive; and

(b) allelic variants of (a).

7. A composition according to claim 6 wherein said one or more polypeptides are further characterized by an absence of N-linked carbohydrate and a presence of O-linked carbohydrate.

8. A composition of mammalian thrombopoietin consisting essentially of one or more polypeptides having an amino acid sequence selected from the group consisting of:

(a) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Arg(198);

(b) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Phe(207); and

(c) allelic variants of (a) and (b).

9. A pharmaceutical composition comprising a thrombopoietin polypeptide in combination with a pharmaceutically acceptable vehicle, wherein said thrombopoietin polypeptide is characterized by an amino acid sequence selected from the group consisting of:

(a) a sequence of amino acids as shown in SEQ ID NO:2 having an amino terminus at Ser(45) and a carboxyl terminus between Ser(208) and Asn(216), inclusive;

(b) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus between Arg(185) and Asn(193), inclusive;

(c) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Arg(198);

(d) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Phe(207); and

(e) allelic variants of (a)-(d).

10. A pharmaceutical composition according to claim 9 wherein said thrombopoietin polypeptide is further characterized by:

a) a peptide backbone mass of 17,490 to 19,045 amu as determined by mass spectrometry;

b) absence of N-linked carbohydrate;

c) presence of O-linked carbohydrate, and said composition is essentially free of thrombopoietin species

having a peptide backbone mass greater than 19,045 amu as determined by mass spectrometry.

11. A pharmaceutical composition according to claim 9 wherein said thrombopoietin polypeptide is characterized by:

(a) an amino acid sequence as shown in SEQ ID NO:4 having an amino-terminus at Ser, residue number 22, and a carboxyl-terminus between Arg, residue number 185, and Asn, residue number 193, inclusive; or

(b) an amino acid sequence which is an allelic variant of (a).

12. A pharmaceutical composition according to claim 9 wherein said thrombopoietin polypeptide is characterized by an amino acid sequence selected from the group consisting of:

(a) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Arg(198);

(b) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Phe(207); and

(c) allelic variants of (a) and (b).

13. A pharmaceutical composition according to claim 9 further comprising erythropoietin.

14. A method of stimulating platelet production in a mammal in need thereof comprising administering to said mammal a therapeutically effective amount of thrombopoietin in combination with a pharmaceutically acceptable vehicle, wherein said thrombopoietin is characterized by an amino acid sequence selected from the group consisting of:

(a) a sequence of amino acids as shown in SEQ ID NO:2 having an amino terminus at Ser(45) and a carboxyl terminus between Ser(208) and Asn(216), inclusive;

(b) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus between Arg(185) and Asn(193), inclusive;

(c) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Arg(198);

(d) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Phe(207); and.

(e) allelic variants of (a)-(d).

15. A method according to claim 14 wherein said thrombopoietin is further characterized by:

a) a peptide backbone mass of 17,490 to 19,405 amu as determined by mass spectrometry;

b) absence of N-linked carbohydrate;

c) presence of O-linked carbohydrate; and

d) essentially free of thrombopoietin species having a peptide backbone mass greater than 19,045 amu as determined by mass spectrometry.

16. A method according to claim 14 wherein said thrombopoietin is characterized by:

(a) an amino acid sequence as shown in SEQ ID NO:4 having an amino-terminus at Ser, residue number 22, and a carboxyl-terminus between Arg, residue number 185, and Asn, residue number 193, inclusive; or

(b) an amino acid sequence which is an allelic variant of (a).

17. A method according to claim 14 wherein said thrombopoietin is characterized by an amino acid sequence selected from the group consisting of:

(a) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Arg(198);

(b) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Phe(207); and

(c) allelic variants of (a) and (b).

18. A method according to claim 14 wherein said mammal is anemic and said method further comprises administering to said mammal an amount of erythropoietin effective to increase erythropoiesis.

19. An isolated DNA molecule encoding a mammalian thrombopoietin characterized by an amino acid sequence selected from the group consisting of:

(a) a sequence of amino acids as shown in SEQ ID NO:2 having an amino terminus at Ser(45) and a carboxyl terminus between Ser(208) and Asn(216), inclusive;

(b) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus between Arg(185) and Asn(193), inclusive;

(c) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Arg(198);

(d) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Phe(207); and

(e) allelic variants of (a)-(d).

20. An isolated DNA molecule according to claim 19 wherein said thrombopoietin is characterized by:

(a) an amino acid sequence as shown in SEQ ID NO:4 from an amino-terminus at Ser, residue number 22, to a carboxyl-terminus between Arg, residue number 185, and Asn, residue number 193, inclusive; or

(b) an amino acid sequence which is an allelic variant of (a).

21. An expression vector comprising a first DNA segment encoding a mammalian thrombopoietin characterized by an amino acid sequence selected from the group consisting of:

(a) a sequence of amino acids as shown in SEQ ID NO:2 having an amino terminus at Ser(45) and a carboxyl terminus between Ser(208) and Asn(216), inclusive;

(b) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus between Arg(185) and Asn(193), inclusive;

(c) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Arg(198);

(d) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Phe(207); and

(e) allelic variants of (a)-(d),

wherein said first DNA segment is operably linked to one or more additional DNA segments that provide for its transcription.

22. An expression vector according to claim 21 wherein said thrombopoietin is characterized by:

(a) an amino acid sequence as shown in SEQ ID NO:4 from an amino-terminus at Ser, residue number 22, to a carboxyl-terminus between Arg, residue number 185, and Asn, residue number 193, inclusive; or

(b) an amino acid sequence which is an allelic variant of (a).

23. A cultured eukaryotic cell containing an expression vector comprising a first DNA segment encoding

a mammalian thrombopoietin characterized by an amino acid sequence selected from the group consisting of:

(a) a sequence of amino acids as shown in SEQ ID NO:2 having an amino terminus at Ser(45) and a carboxyl terminus between Ser(208) and Asn(216), inclusive;

(b) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus between Arg(185) and Asn(193), inclusive;

(c) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Arg(198);

(d) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Phe(207); and

(e) allelic varaints of (a)-(d),

wherein said first DNA segment is operably linked to one or more additional DNA segments that provide for its transcription.

24. A cultured eukaryotic cell according to claim 23 wherein said thrombopoietin is characterized by:

(a) an amino acid sequence as shown in SEQ ID NO:4 from an amino-terminus at Ser, residue number 22, to a carboxyl-terminus between Arg, residue number 185, and Asn, residue number 193, inclusive; or

(b) an amino acid sequence which is an allelic variant of (a).

25. A cultured eukaryotic cell according to claim 23 which is a yeast cell.

26. A cultured eukaryotic cell according to claim 23 which is a mammalian cell.

27. A cultured eukaryotic cell according to claim 26 which is a rodent cell.

28. A method for producing a mammalian thrombopoietin comprising the steps of:

culturing a eukaryotic cell containing an expression vector comprising a first DNA segment encoding a mammalian thrombopoietin characterized by an amino acid sequence selected from the group consisting of:

(a) a sequence of amino acids as shown in SEQ ID NO:2 having an amino terminus at Ser(45) and a carboxyl terminus between Ser(208) and Asn(216), inclusive;

(b) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus between Arg(185) and Asn(193), inclusive;

(c) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Arg(198);

(d) a sequence of amino acids as shown in SEQ ID NO:4 having an amino terminus at Ser(22) and a carboxyl terminus at Phe(207); and

(e) allelic variants of (a)-(d),

wherein said first DNA segment is operably linked to a second DNA segment encoding a secretory peptide and one or more additional DNA segments that provide for transcription of the first and second DNA segments, and wherein said cell expresses said first and second DNA segments and said thrombopoietin is secreted from the cell; and

selectively recovering said thrombopoietin.

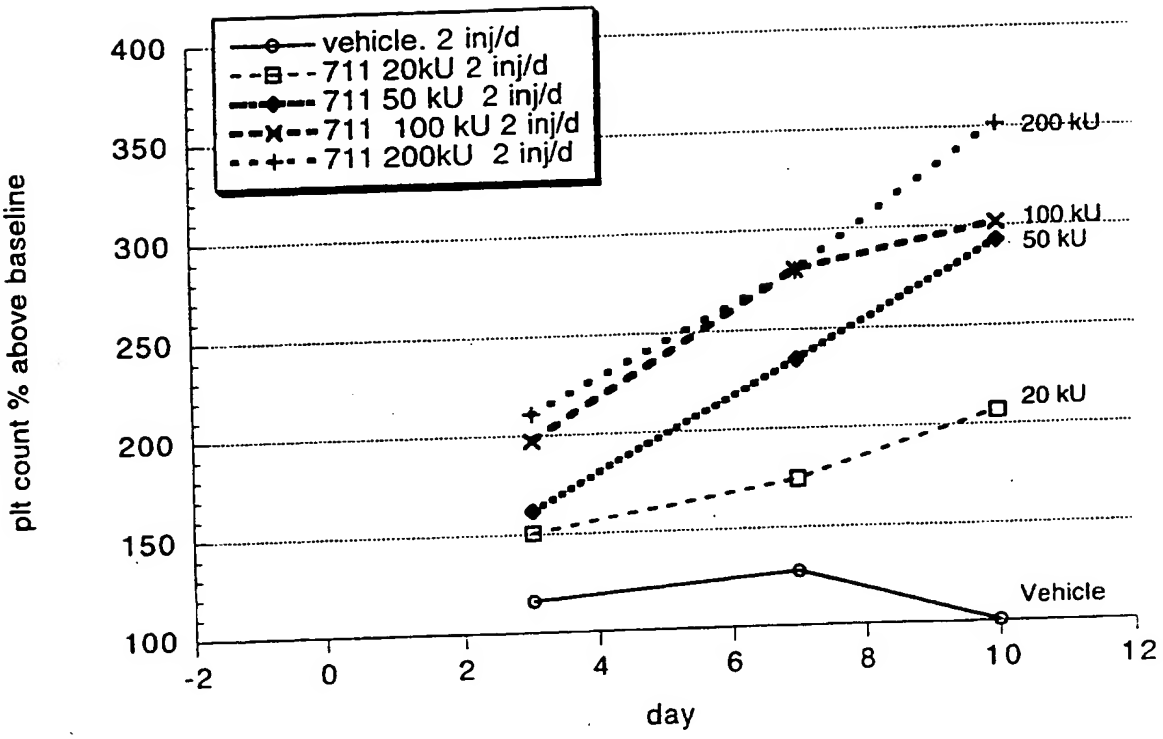

vehicle. 2 inj/d
711 20kU 2 inj/d
711 50 kU 2 inj/d
711 100 kU 2 inj/d
711 200kU 2 inj/d
plt count % above baseline
400
350
300
250
200
150
100
-2
0
2
4
6
8
10
12
day
200 kU
100 kU
50 kU
20 kU
Vehicle

Fig. 1

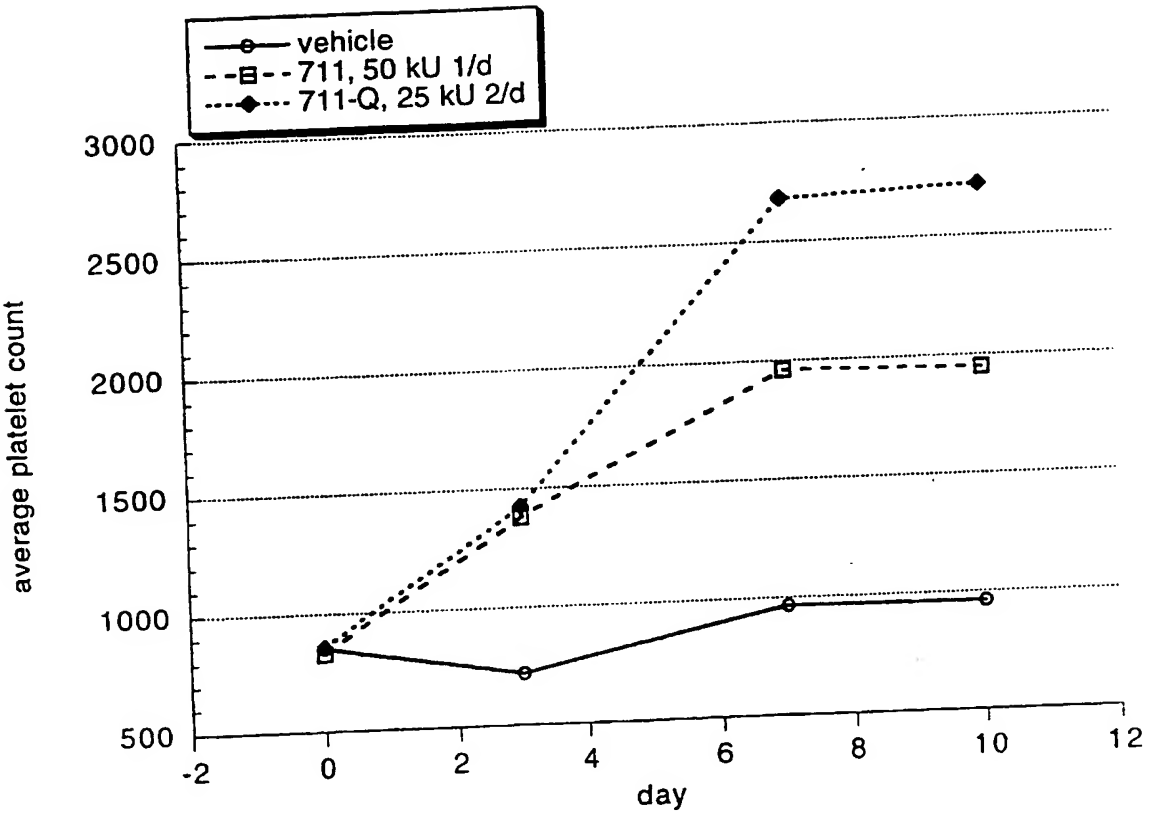

vehicle
711, 50 kU 1/d
711-Q, 25 kU 2/d
average platelet count
3000
2500
2000
1500
1000
500
-2
0
2
4
6
8
10
12
day

Fig. 2

INTERNATIONAL SEARCH REPORT

Interr nal Application No
PCT/US 95/14929

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/19 C12N15/79 C07K14/52 A61K38/19 C12N1/19
C12N5/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	NATURE, vol. 369, no. 6481, 16 June 1994 LONDON GB, pages 565-568, SI LOK ET AL. 'Cloning and expression of murine thrombopoietin cDNA and stimulation of platelet production in vivo' cited in the application see the whole document --- -/--	1,6-9, 14,19, 21,23, 26-28

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 March 1996	27.03.96

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Montero Lopez, B

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	EP,A,0 675 201 (AMGEN INC) 4 October 1995 see page 3, line 3 - line 7 see page 7, line 34 - line 40 see page 8, line 13 - line 15 see page 8, line 30 - line 50 see page 9, line 28 - line 39 see page 11, line 16 - page 12, line 16 see page 14, line 16 - line 30 see page 14, line 42 - page 15, line 30 see page 15, line 41 - line 57 ---	1,4-9, 11-14, 16-28
T	J. BIOCHEM. (TOKYO) (1995), 118(1), 229-36 CODEN: JOBIAO;ISSN: 0021-924X, 1995 KATO, TAKASHI ET AL 'Purification and characterization of thrombopoietin' -----	

1

INTERNATIONAL SEARCH REPORT

ternational application No.
PCT/US95/14929

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 14-18
because they relate to subject matter not required to be searched by this Authority, namely:
Remark: Although these claims are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound.

2. [ ] Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. [ ] As all searchable claims could be searches without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter nal Application No
PCT/US 95/14929

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP-A-0675201	04-10-95	AU-B-	2230895	23-10-95
		EP-A-	0690127	03-01-96
		WO-A-	9526746	12-10-95